DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Dissertation Committee for Jessica Ann White-Phillip Certifies that this is the approved version of the following dissertation:

Glycosylation Reactions in Secondary Metabolism: Glycosylation Events in *C*-mannosylation and the Biosynthesis of Kijanimicin

Committee:

Hung-wen Liu, Supervisor

Karen Browning

Eric Anslyn

Sean Kerwin

Zhiwen Zhang

Glycosylation Reactions in Secondary Metabolism: Glycosylation Events in *C*-mannosylation and the Biosynthesis of Kijanimicin

by

Jessica Ann White-Phillip, B. S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2010

Dedication

To my family- for their enduring inspiration, encouragement, and support.

Acknowledgements

There are so many people I wish to thank. I will start with those who helped me in the lab. First, thanks to all members of the Liu Group, present and former for being such a great group of people to work with. I always found someone (or someone's dissertation) to help me learn a new technique or procedure. Thanks to all former, present and future members of the "Sugar Biosynthesis and Natural Products Subgroup". Thank you for teaching me and helping me learn how to teach. Thanks to Drs. Svetlana Borisova, Lin Hong, Hua Zhang, and Haruko Takahashi that taught me everything I know about sugar biosynthesis, sugar purification, glycosyltransferases and proper HPLC techniques. Thanks to Dr. Charles (Chad) Melançon for helpful discussions regarding experimental conditions, writing critiques, literature sources and demonstrating to me that even at an advanced age one can always acquire a "pesky" older brother. Thank you to Dr. Hyung-jin Kwon, (a.k.a. Captain Stuff) who taught me everything I needed to know about working in *Streptomyces*, and informing me of the fact that "Any idiot can get a protein in *E. coli*, but it takes a real scientist to get one in *lividans*." Well, I managed to be a "real scientist", thank you. Thanks to Drs. Steven Mansoorabadi and Mark Ruszczycky, (the two post-docs with the hardest last names to spell) for helpful discussions and reading

chapters of this dissertation. Thanks to Drs. Zhihua Tao and Ying Zhou, my "older sisters" and office mates in the lab, who never hesitated to make our office a bright, cheery place to be. Thanks to my "younger sisters" and office mates Chia-I Lin and Mei-lan Wu, who continue to keep our office full of smiles and laughter. Thanks to Dr. Wendy Marriner for reading parts of my dissertation, and teaching me everything I know about organic chemistry. Thank you for the coffee breaks, thanksgiving pot-lucks, and sharing my frustrations, but mainly- thank you for your friendship. Thanks to my undergraduates: Bryant "Whoops!" Chambers, Joel "Joe" Robertson, Esther "Mei-mei" Kwong, Nara "Noodles" Shin, Haibi (Daniel) "Daniel-san" Cai, Sun "sunshine" Cho, Michael "Mike Wazowski" Scott, and last but definitely not least, Anthony/Tony "#9" Romo. All of you taught me so much about patience, encouragement, direction, explanation, record-keeping, learning and guidance in your own way. I learned as much from you as (I hope) you learned from me.

I am truly grateful to my P. I. Dr. Hung-wen "Ben" Liu for his leadership, encouragement and belief in me, and Dr. Yung-nan Liu for her guidance, technical expertise, generosity and friendship. Both of you encouraged me every step of my journey, allowing me to choose a risky project and letting me apply my tenacity. Ben, you never watched over my shoulder to guide me, only asked the right questions to help me move in the direction. Yung-nan, you were always there for me. Whenever I needed anything, I knew I could come to you, whether it was help with a research topic or a technique, or just someone to talk to about a problem. You are a true asset to the Liu lab, and I greatly value everything you give to the students and post-docs in the research group. Thank you sincerely; the opportunity to work in the Liu lab has been a privilege.

I must also thank my family. I must thank my parents for my college education, and raising me to understand that I can attain what I choose, as long as I am willing to work for my goals. Thank you to my father, who has grown to be my friend. Dad and Katie, thank you for your advice and guidance throughout my life especially in graduate school, and for your help and support during the hard times. Mom, thank you for your encouragement, and for believing in me (and thank you most of all for hemming my pants before my oral examinations so I don't have to use scotch tape). If I tried to thank you for everything you have done for me, my dissertation would be 1000 pages long and half of that would be acknowledgements. Thanks to my "Aunties" (in alphabetical order): Jane, Katy and Mary for advice on life, living and the pursuit of happiness, as well as (when within 50 feet of each other) being a "party in a box" while at all times fully prepared for every potential catastrophe no matter how large or small, and more importantly, assisting me to learn how to do the same. Thanks to my Uncle Francis for his enduring peace and understanding, who more often than not had to put up with my aunties and me when we were all together. Thanks to my grandmother, Anna Bailey (1914-2008) for her unconditional love, graciousness and generosity to me. Thanks to my father-in-law Charles C. Phillip, whose hunting prowess and woodcraft will never be equaled by another man alive in my eyes, nor will his stubbornness, practicality and kind heart. Thank you Charlie for all you did for me, and you were right about Sarah Palin. (1943-2008) Thanks to my sister, Logan for being my friend and "growing up" with me.

Thank you to my former professors Dr. Alice Sessions and Dr. Rudy Bauss, who helped me find in myself a love of learning and a love of teaching. Thank you to Laura Benton, for allowing me to teach her, and be taught by her.

Laura, you are such a wonderful friend, student and teacher, I am grateful for your help and friendship through the most stressful times. Thank you Dr. von Eschenbach, for your inspirational words, career advice and friendly assistance.

Thank you to my wonderful husband, who put up with long nights, early mornings and schedules than changed week to week. William, you are more of a man than any woman could ever hope of having. You have supported me, encouraged me, and accompanied me on every step of my journey. William, you have gone beyond the call of duty, listening to me practice my science talks with your artist ears, tried to help me solve biology problems with your artist brain, and managed the various issues in life so that I didn't have to take care of them. Thank you for cooking dinner every night for seven years, packing my lunches, and waking me up every morning with hot caffeinated beverages. Thank you for all the things you did and still do for me every day of our lives together. I would not have made it this far without you. (The cat also thanks you for remembering to feed her when I didn't, and has stepped on the keyboard and typed this mark to tell you so ;p;.)

Finally, thank you to all those who are not named here explicitly, but helped in some way on one day or another. Thanks to all those who have guided me, whether you knew it or not, I was listening. Thanks to all those who rejoiced in my successes, and encouraged me in my defeats. Thanks to all those who allowed me to make mistakes and try again. As in all cases, I must end with the appropriate: the successes here are theirs- those of my family and friends, my teachers and students, all failures are mine and mine alone.

Glycosylation Reactions in Secondary Metabolism: Glycosylation Events in *C*-mannosylation and the Biosynthesis of Kijanimicin

Publication No._____

Jessica Ann White-Phillip Ph. D.

The University of Texas at Austin, 2010

Supervisor: Hung-wen Liu

In this work, we examine two disparate aspects of glycosylation. The first project involves the elucidation of the glycosylation of the novel tetronolide natural product, kijanimicin. The biosynthesis of the deoxysugar TDP-L-digitoxose from the kijanimicin natural product pathway was achieved *in vitro*. The genes were identified from the cluster, cloned, expressed and the products were purified. Activity was demonstrated for the novel enzymes and the pathway was reconstructed *in vivo* using *Streptomyces lividans*. These strains of *S. lividans* were used to examine kijanimicin glycosyltransferase activity. We were able to demonstrate activity for 3 of 4 digitoxosyltransferases in the biosynthetic pathway and propose a biosynthetic scheme by which the tetrasaccharide chain is formed. We identified two putative glycosidases with novel folds, and one

glycosyltransferase that appears to have unprecedented activity, attaching 2 if not 3 sugars in sequence.

In the second portion of this work, we attempted to identify the eukaryotic *C*-mannosyltransferase enzyme and demonstrate its activity *in vitro* and *in vivo*. Here, we describe our efforts to identify the CMT. Through *in silico* analysis, putative *C*-mannosyltransferase genes were identified. These genes were expressed in *E. coli* and *S. cerevisiae*, however gene expression was apparently toxic to *E. coli*. *S. cerevisiae* expression was acceptable, but extraction proved to be somewhat problematic. We describe our efforts to develop a CMT assay for use *in vitro* by expressing the putative CMT in insect cells, which was much more promising. We also attempted to knock down the putative CMT genes using shRNA, which demonstrated that the genes of unknown function that were identified were essential for cellular viability.

This work has contributed to the fields of both *C*-mannosylation and natural product glycosylation. We have elucidated the biosynthetic pathway of a novel deoxysugar, and identified potentially valuable tools for glycoengineering including a glycosyltransferase that appears to exhibit novel polymeric activity, as well as identifying two glycosyltransferase proteins that are apparent glycosidases. Our attempts to identify the CMT provided valuable insight into the future development of a *C*-mannosylation assay, and we have identified several promising protein candidates that are apparently essential for *H. sapiens* cellular viability.

Table of Contents

List of Tables

List of Figures

List of Illustrations

1 An Overview of Eukaryotic and Prokaryotic Glycosylation

Glycosylation is the most common and most diverse post-translational modification that newly synthesized proteins may undergo.[1] Covalent linkages between proteins and carbohydrates have been found in all organisms, including archaea, prokaryotes and eukaryotes.[2] It has been argued that glycosyltransfer reactions are, on quantitative terms, the most important biological transformations on Earth, since they account for the biosynthesis and hydrolysis of the bulk of biomass.[3] Thus, one may also argue that the most important enzymes are those involved in glycosylation reactions, including sugar donor biosynthesis, glycosyltransferases or enzymes which catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules to form glycosidic bonds, and glycosidases, which catalyze the removal of sugar moieties from glycosylated acceptor molecules. Notably, the biosynthesis of disaccharides, oligosaccharides and polysaccharides involves the action of literally hundreds of different glycosyltransferases.[4]

Eukaryotic glycosylation encompasses multiple forms of protein post-translational modifications including *N*, *O* and the newly discovered *C*-glycosylation. Glycosylation in prokaryotes is more limited, with protein modifications involving simple *N*- and *O*-glycosylation only, although myriad secondary metabolites and other natural products are highly decorated with diverse glycoforms. In this Chapter, a brief overview of prokaryotic and eukaryotic glycosylation will be presented, with an emphasis on the respective similarities and differences between the systems. To demonstrate the

importance of all forms of glycosylation in both prokaryotes and eukaryotes, model or unique pathways will be described including S-layer glycans involving *N*- and *O*-glycosylation, lipopolysaccharide biosynthesis (prokaryotic *O*-glycosylation), prokaryotic *N*-glycosylation, natural product glycosylation in prokaryotes, which includes *C*-, *O*- and *N*-linkages, several forms of eukaryotic *O*-glycosylation, eukaryotic *N*-glycosylation, and finally *C*-mannosylation, which is found only in eukaryotes.

1.1 ROLE OF PROKARYOTIC GLYCOSYLATION

Prokaryotes possess an extreme variety of carbohydrates as cell surface constituents. These carbohydrates are either attached to lipids or are part of elongated glycan chains. Glycoconjugates that are linked to proteins or peptides generated by ribosomes had never been reported prior to the identification of an S-layer glycopeptide found in an archaeal halophile.[5] While this S-layer glycan was proposed to play a role in maintaining cell shape, little research followed as S-layer glycans were not considered to be medically relevant.[5] Prokaryotic glycosylation regained attention when glycosylated proteins were identified in bacterial pathogens. In fact, most glycoproteins in Gram-negative bacteria are associated with virulence factors of medically significant pathogens.[6, 7] Since the discovery of prokaryotic protein glycosylation, surface protein glycosylation has been linked to important steps in pathogenesis. Specifically, bacterial protein glycosylation may have specific functions in infection and pathogenesis, while interfering with inflammatory immune responses.[7] It has been suggested that these glycans serve multiple functions in infectious pathogens. Prokaryotic protein glycosylation has been suggested to include the maintenance of protein conformation,

surface recognition, increasing resistance against proteolytic digestion, enzymatic activity, cell adhesion and immune evasion, and has been shown in some pathogens to function in intercellular interactions such as adhering to host cells.[7] The relationship between pilin glycosylation and cellular adhesion clearly demonstrates the link between glycosylation and pathogenesis in bacterial infection.[7] It has also been suggested that another role of prokaryotic glycosylation is to modulate immune response by interacting with antibodies from the host.[7] Modulation of immune response could result from altering pathogenic glycoforms, which would in turn affect host antibody production, thus influencing the infection.[7]

1.1.1 Secondary Cell Wall Polymers

S-layer or "surface layer" glycosylation is the oldest known and perhaps the best studied prokaryotic glycosylation, although glycosylated enzymes, antigens and cell-envelope components have also been identified.[2] The first S-layer glycoprotein was found in halophile archaeon *Halobacterium halobium*.[8, 9] After the initial identification of S-layer glycoproteins, this type of glycosylation was also found in prokaryotic organisms, mainly in low G+C Gram-positive bacteria.[2] Since the identification of S-layer glycoproteins, linkage regions has been identified and S-layer proteins have been cloned and sequenced.[2] in addition, some information about the glycan biosynthetic pathways is also available.[2]

In bacteria and archaea, the outermost cell envelope layers are 5-20 nm thick and cover the cell surface completely.[2] Subunits composing the cell envelope are 40-200 kDa and are arranged in lattice structures spaced 2.5-35 nm apart. [2] These glycans protrude 30-40 nm from the cell surface to create a polysaccharide coat similar to that of LPS

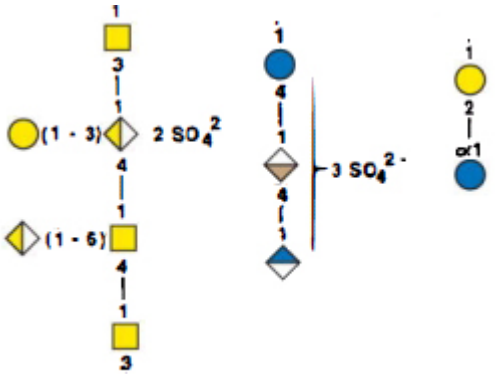
antigens of Gram-negative bacteria.[2] S-layer glycans are planar assemblies of weakly acidic glycoproteins found in archaea and Gram-positive organisms with isoelectric points ranging from pH 4 to pH 6. [2, 10, 11] Proteins in S-layers are frequently post-translationally modified by glycosylation, phosphorylation, and protein termini (*N*- and *C*-) are trimmed.[2] S-layer glycans are comparable to the lipopolysaccharide (LPS) coat of Gram-negative bacteria, however the difference is how the glycans are anchored to the cell wall.[2, 12] In bacterial S-layers, the glycan is anchored through a core structure, much like LPS are anchored via lipid A, while in archaea the S-layer anchor is a polypeptide, linked by *N*- or *O*-glycosylation.[12] The existence of a core-oligosaccharide connecting the bacterial S-layer polysaccharide and the protein portion has been experimentally determined. [13-15] This data leads to the deduction of a tripartite architecture for these cell-surface macromolecules.[2] The overall structure consists of long, variable *O*-antigen-like carbohydrate chains and the oligosaccharide core is glycosidically linked to the S-layer polypeptide. [2]

In archaea, *N*-linkages predominate whereas in bacteria, *O*-linkages such as β-Gal→Tyr, β-Glc→Tyr or β-GalNAc→Thr and β-GalNAc→Ser/Thr are most abundant.[2] Bacterial S-layer glycans may also be attached through a rare Rha→Asn *N*-linkage, although *O*- and *N*-linkages are never found on the same protein. In contrast, archaeal S-layers contain *N*- and *O*-linkages on the same protein, including *N*-linkages of Glc→Asn, GalNac→Asn, Rha→Asn and *O*-linkages to Thr.[2] Other differences between bacterial and archaeal S-layer glycans are the number of glycans attached to the protomer and the polymer size. Bacteria have fewer attachment sites (2-6), and longer polymers (20-50) in repeating units of 2-6 saccharides.[2] In contrast, archaea generally have a higher number of attachment sites (3-25 per protein), shorter polymers (up to 10 saccharides), and have no repeating saccharide units.[2, 16] Generally, a single type of carbohydrate chain is

generally found per organism, although archaea may exhibit up to 3 different types of glycans on one protomer. These glycans may vary in chain length, and glycosylation sites on the protein including amino acid acceptor sequence variation.[2]

In bacteria, these polymers can be linear or branched structures, and include a wide range of hexoses, deoxysugars, *N*-acetylated aminosugars, uronic acids, which can also be derivatized with phosphate, sulfate and glycerol groups. [17-19] It should be noted that several sugars are found exclusively in the LPS of Gram-negative bacteria, such as D-rhamnose 3-*N*-acetyl- D-fucosamine, D-glycero-D-manno-heptose and others, which have been isolated from bacterial S-layer glycans.[20, 21]

1.1 Examples of S-Layer Glycans from Archaea



In bacteria and archaea, *N*-glycans are found to link to the consensus sequence N-X-S/T where X is any amino acid except proline.[2] It should be noted that S-layer glycosylation is not restricted to conventional sequences: i.e. e., asparagine for *N*-glycosidic linkages and threonine, serine and hydroxyproline or hydroxylysine for *O*-

glycosidic linkages, but also include tyrosine as potential glycosylation sites.[22] *O*-glycosylation may occur in Pro rich domains as in eukaryotes, but may also be found on consensus sequences such as D-S, D-T-T, or V-Y. [22]

1.1.1.1 Biosynthesis of S-Layer Glycans

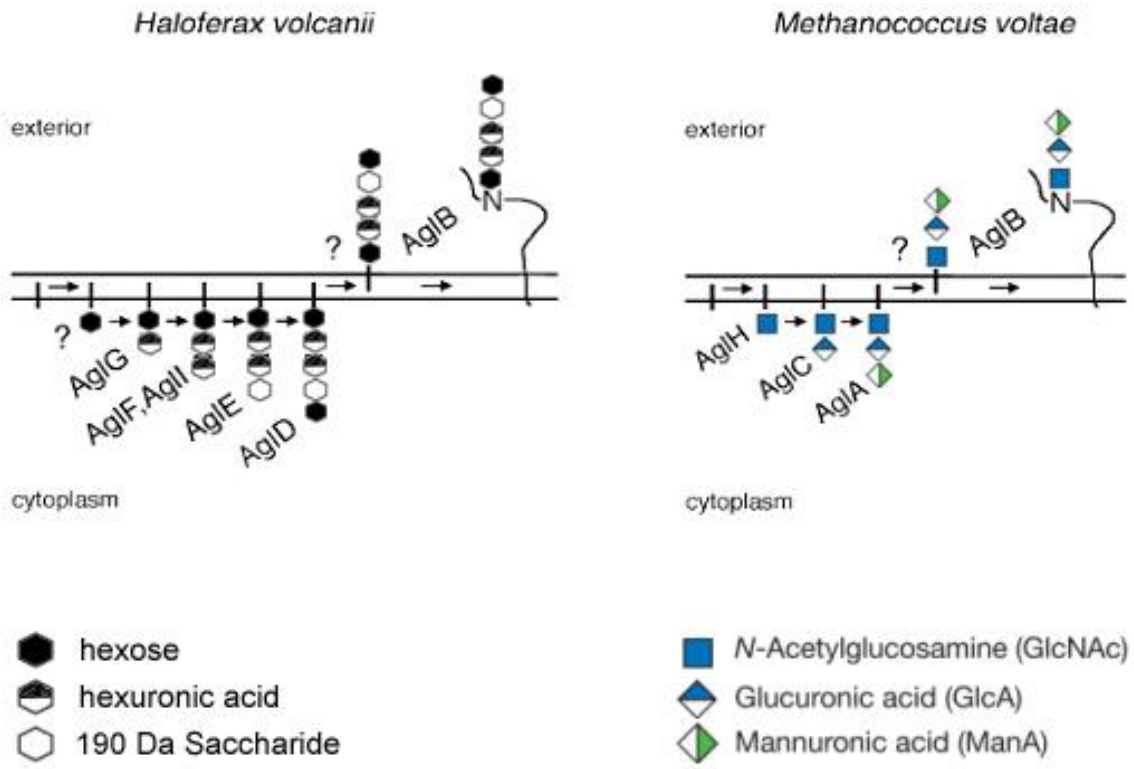
Forty S-layer glycoprotein structures have been partially or fully elucidated, but much less is known with respect to the biosynthesis of these glycans.[23] Archaeal *O*-glycosylation pathways are almost completely unknown.[23] It was previously assumed that archaeal pathways follow the general format of *N*-glycosylation in other organisms such that sugars are assembled into a polymer on a lipid carrier. The lipid carrier will be translocated across the cellular membrane followed by protein glycosylation. Archaea are known to use both dolichol phosphate and undecaprenyls for lipid anchors in glycosylation.[5, 23, 24] As in other forms of *N*-glycosylation, the glycans are assembled on the lipid anchor in archaea.[25] Once the oligosaccharide is complete, it is translocated across the membrane by unidentified cellular machinery.[25] A few enzymes involved in these biosynthetic pathways have been characterized.[23] But due to the diversity of glycan structures, S-layer pathways from each organism must be individually characterized.[23]

Much research remains to be done in elucidating the biosynthetic pathways of nucleotidyl activated sugar substrates in archaea, although studies in methanogens and *Thermococcales* demonstrated that these pathways generally follow the bacterial biosynthetic routes.[26] Some glycosyltransferase genes and their products in the archaeal S-layer pathway have been characterized. In *M. voltae*, AglC and AglA add the penultimate and the terminal 2,3-diacetamido-2,3-dideoxy-β-glucuronic acid moieties, respectively, while AglH, a homolog of the yeast Alg7 protein, is involved in addition of

the Asn-bound linking sugar, GlcNAc.[27, 28] In *H. volcanii*, AglD, AglE, AglF, AglG, and AglI transfer the second to the fifth sugar subunits of the partially characterized pentasaccharide linked to the S-layer glycoprotein.[25, 29, 30]

More work remains to be done in this area to elucidate these biosynthetic pathways. However, these examples cited here serve to demonstrate current evidence for the stepwise addition of sugars by dedicated glycosyltransferases to the lipid carrier for the production of a lipid-linked oligosaccharide.

1.2 Archaeal S-Layer Pathways from *H. volcanii* and *M. voltae*



1.1.1.2 Archaeal S-Layer N-Glycosyltransferase

In both *M. voltae* and *H. volcanii*, aglB encodes the archaeal OST, which comprises a single subunit homologous to the core STT3 subunit of the multimeric eukaryotic OT complex.[23] As in eukaryotes, the canonical N-X-S/T motif is a sufficient condition governing the site of glycosylation in archaea.[25] In archaea, however, the replacement of Ser with Val, Leu, or Asn does not prevent glycosylation, which suggests the presence of an *N*-glycosyltransferase enzyme which is more promiscuous than those found in bacteria or eukaryotes.[31] However, the N-glycan profile of the *H. salinarum* S-layer glycoprotein reveals the presence of two different oligosaccharides, each linked to selected Asn residues through either glucose or GlcNAc subunits, which suggests that the archaeal OT may tolerate a range of sugar substrates, unlike its eukaryotic or bacterial OT homologs.[26]

A crystal structure of the *C*-terminal domain of the aglB enzyme was solved to 2.7 Å. This data provided important structural information regarding the archaeal OST and could be related to the bacterial and eukaryotic OT proteins.[32] The crystal structure showed a central core domain consisting of α-helices that includes the WWDYG catalytic motif, with β strands surrounding the central core.[23] Alignments of Stt3 (OT) homologs from eukaryotes, bacteria and archaea enables the discovery of a conserved DXXK motif.[23] According to the crystal structure, the conserved Asp and Lys residues are located on a helical face near the catalytic motif. Thus, it was proposed that the DXXK motif may interact with the pyrophosphate moiety of the lipid carrier.[23] Subsequent site-directed mutagenesis in yeast provided experimental support for the catalytic involvement of the DXXK motif in *N*-glycosylation.[23] In the archaeal AglB, DWQMASTDAK is found instead of the DXXK motif, which suggests that the disruption of the widely conserved DXXK motif may correlate with the use of dolichol

mono- rather than pyrophosphate as the oligosaccharide-charged lipid carrier in these organisms.[26]

1.1.1.3 Role of S-Layer Glycans

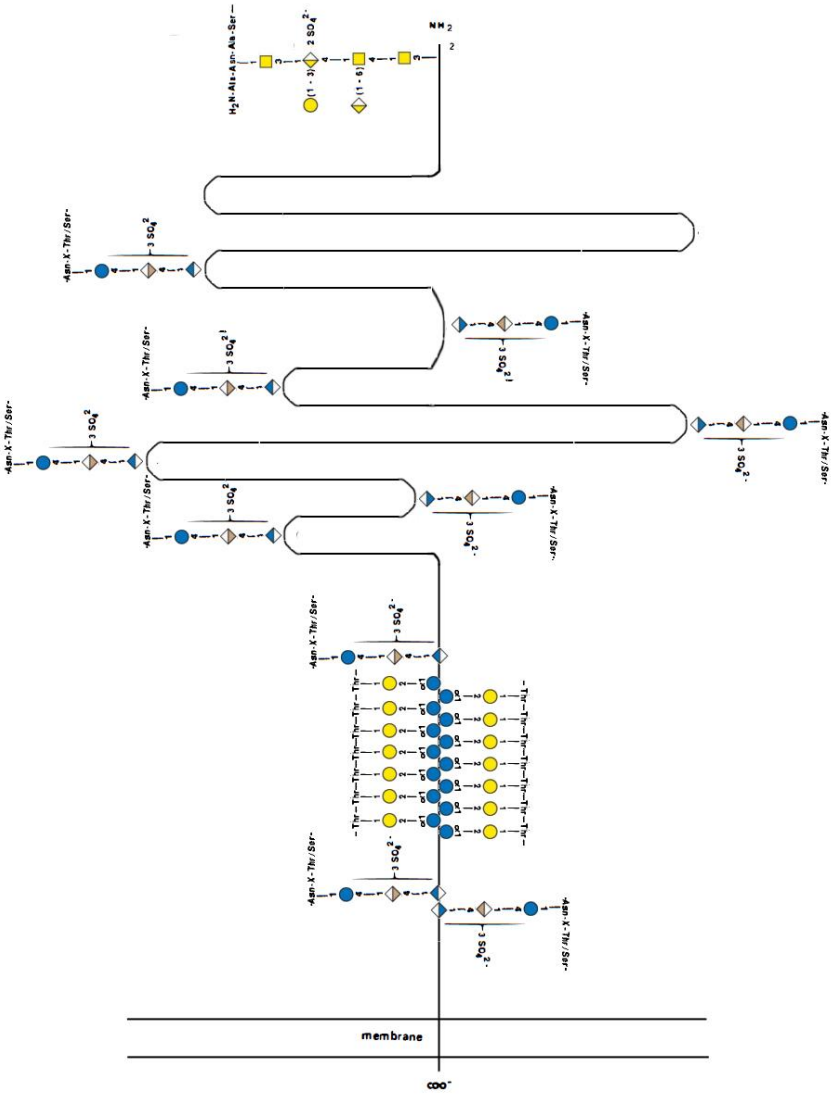
When aglB is removed by mutation from *M. voltae* and *H. volcanii*, the cells remain viable. Thus, *N*-glycosylation does not appear to be essential for cell survival in archaea, unlike eukaryotes.[29] However, the inability to perform *N*-glycosylation renders archaea unable to grow at elevated salt concentrations and to display unstructured and defective S-layers.[29] In *M. voltae*, lack of *N*-glycosylation results in the reduction of flagella (overall) and the cells also display motility defects. [29]

The function of prokaryotic glycans has been experimentally demonstrated in several cases, and those in the S-layer of the organism are the most well studied.[5] S-layer glycans have been shown to provide a framework for cell shape, aid in cell division process, involve cell adhesion, surface recognition, protective coats, act as molecular sieves or molecule and ion traps, and provide adhesion for enzymes.[2] Taken together, these data suggest that while not essential, *N*-glycosylation provides adaptive benefit for archaeal cells that may allow these organisms to adapt and thrive in extreme environments.[26]

Initially, there a lack of interest in S-layer glycans until they were identified in pathogens, and these glycans received even more attention when protein glycosylation was correlated with prokaryotic pathogenicity.[7, 33] Early reports of S-layer involvement in infectious capacity of organisms described a glycoenzyme and a platelet aggregation-associated glycoprotein of *Streptococcus sp*. strains.[34, 35] Further reports contained detailed structural analysis of the 45-kDa soluble antigen of *Mycobacterium tuberculosis,*

which was found to contain short α-linked mannose chains in *O*-linkage to Thr residues near the *N*- and *C*- termini.[36] These early reports led to the investigation of other glycosylated cell surface appendages, which will be discussed in a later section.

1.3 Multiple S-layer Glycans May Coexist on One Protomer



1.1.2 BACTERIAL PROTEIN GLYCOSYLATION

There exists in prokaryotes a greater diversity of glycan composition and glycosylation sequences than eukaryotes.[22] There are two possible pathways for protein glycosylation in Gram-negative bacteria. One pathway requires supplementation from the lipopolysaccharide or lipooligosaccharide (LPS, LOS) biosynthesis pathways for sources of activated carbohydrate precursors, the necessary biosynthetic enzymes and glycosyltransferases as in the case in *E. coli* and *Pseudomonas O*-glycosylation.[37, 38] The second methodology employed by bacteria to glycosylate proteins involves the evolution or acquisition of separate gene clusters that encode all genes required for precursor and glycan synthesis as well as glycosyltransferases for attachment.[7, 39] It is the second methodology employed by *C. jejuni* and other organisms that use the pgl locus to accomplish protein *N*-glycosylation.

1.1.2.1 N-glycosylation "Pgl" in C. jejuni

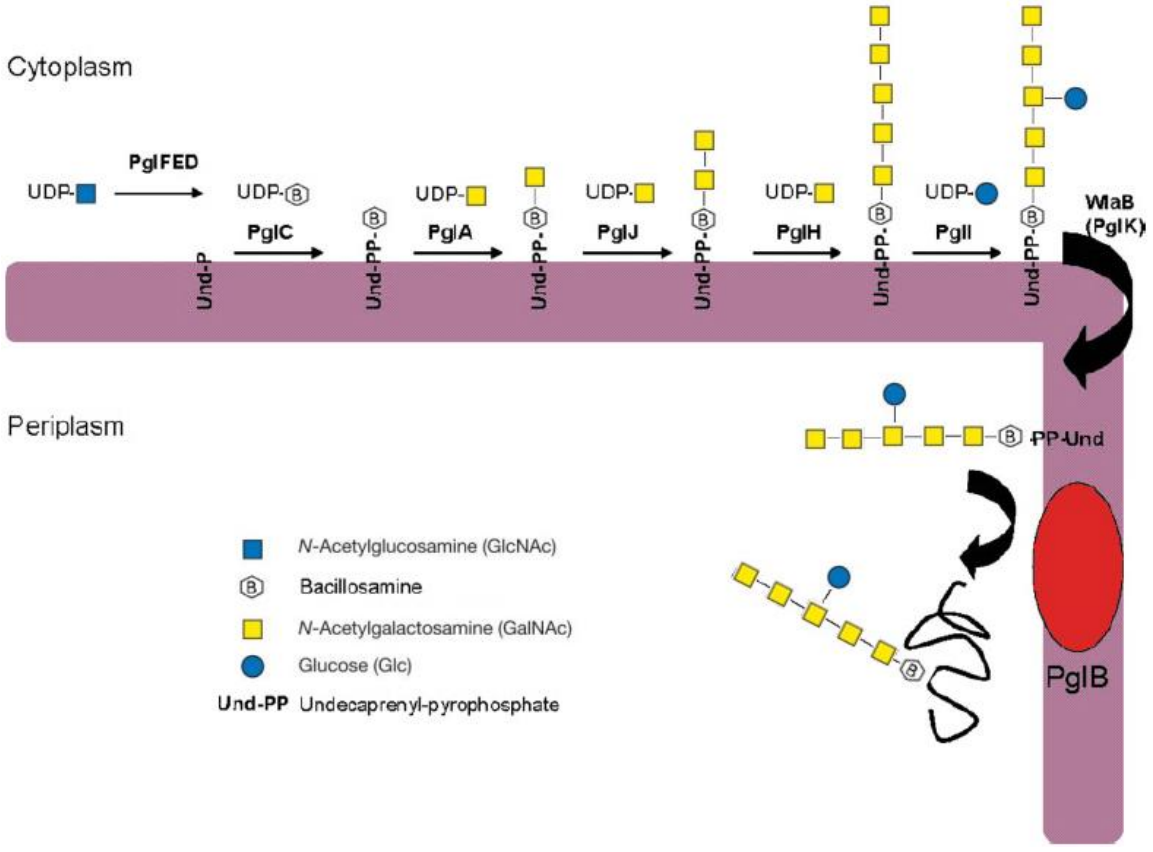
In eukaryotic proteins, *N*-glycans are found at an expanded consensus sequence N-X-S/T where X is any amino acid except proline. However, an expanded D/E-Z-NX-S/T sequon motif, where Z and X are not prolines, is required for *C. jejuni N*-glycosylation. [22, 40] Another notable difference is in the structure of prokaryotic glycans. The latter differ from those of eukaryotes in their lack of antennary structures.[22] The biosynthesis of glycans in prokaryotes is also different. In eukaryotes, most *N*- and *O*-glycosylation occurs in the endoplasmic reticulum and Golgi apparatus. However, there are no membrane bound organelles in most prokaryotes. Thus, glycan biosynthesis must occur in the cytosol of the organism, employing the cell membrane when biosynthesis requires lipid-bound substrates.[22] The lipid carrier employed by bacterial glycosylation machinery is also different from eukaryotes. Bacterial systems utilize undecaprenyl

pyrophosphate, while eukaryotic systems use dolichol pyrophosphate.[41] Nevertheless, overall biosynthetic pathways remain quite similar, as lipid bound and nucleotidyl activated sugars are used as substrates in oligosaccharide chain assembly. [22]

The most well-characterized prokaryotic *N*-glycosylation pathway (figure 1.1) is found in *Campylobacter jejuni*, a human gut mucosal pathogen that is a main cause of bacterial gastroenteritis worldwide.[24] The proposed functions of proteins encoded by the pgl locus in *C. jejuni* are based on previous studies of this organism and on published reports of similar protein sequences in *Neisseria meningitidis* and other organisms. The first step in prokaryotic *N*-glycan biosynthesis is the attachment of an *N*-acetylhexosamine (HexNAc) residue to a lipid carrier, undecaprenyl pyrophosphate by PglC.[39] The HexNAc is subsequently converted to bacillosamine on the lipid by the sequential actions of PglF (dehydratase), PglE (aminotransferase) and PglD (acetyltransferase).[39] In *N. meningitidis*, PglA is the glycosyltransferase responsible for the transfer of α-1,3-linked galactose to bacillosamine and possibly to GlcNAc, therefore, it is proposed that PglA in *C. jejuni* adds the α-1,3-linked GalNAc to bacillosamine.[39, 42] PglH and PglJ have similar amino acid sequences and it has been demonstrated that PglH is an α-1,4-GalNAc transferase.[39] In addition, mutation experiments performed on PglH and PglJ demonstrated that they are both functional.[43] Thus, both proteins could be involved in transferring the next four α-1,4-linked GalNAc residues. The remaining single-branched glucose residue could be added in sequence or after the GalNAc polymer is assembled, and the remaining glycosyltransferase, PglI, is proposed to be responsible for the addition of this sugar residue.[39] Both WlaJ and pglG are omitted from this glycosylation model. The former was not considered due to its unknown function and the latter was shown to be non-essential.[39, 44, 45]

In the current proposal, the assembled heptasaccharide is flipped across the inner membrane by the ABC transporter, WlaB.[39] This is analogous to the eukaryotic pathway where the assembled sugars are flipped into the lumen of the endoplasmic reticulum (Figure 1.4).[46] It should be noted that at this stage of biosynthesis in both the bacterial and eukaryotic systems, the potential *N*-glycan is a lipid-linked heptasaccharide containing one branched sugar.[39, 46] In eukaryotes, the heptasaccharide is further processed to a tetradecasaccharide, which is then transferred to proteins by the oligosaccharyltransferase protein complex, including the PglB homologue, Stt3.[46]

1.4 The pgl *N*-Glycosylation Pathway of *C. jejuni*



1.1.2.2 PglB- the Model Prokaryotic N-glycosyltransferase

The key glycosyltransferase enzyme in the pgl gene locus is PglB. This protein shows significant amino acid similarity to the staurosporine- and temperature-sensitive yeast protein 3 (Stt3) subunit of the *N*-linked oligosaccharyltransferase (OT) complex of *Saccharomyces cerevisiae*.[45] Stt3 has recently been shown to play a central role as the glycosyltransferase or as the catalytic subunit of the nine-member OT complex.[47] Stt3 homologs are found in several eukaryal and archaeal species and are characteristically large proteins (700–970 amino acids) with 10–12 potential membrane-spanning regions and a conserved C-terminal catalytic domain (WWDYG) similar to PglB.[44, 47] In *C. jejuni*, the OT homolog PglB appears to be the only protein required for glycosyltransfer of the glycan to the common *N*-linked sequon.[39] Another deviation from eukaryotic *N*-glycosylation systems is a lack of evidence for further processing of the heptasaccharide.[39] In contrast to the eukaryotic system, in which *N*-glycosylation targets proteins to folding chaperones in the ER, the *N*-glycosylation system in prokaryotes appears to modify proteins that have already been folded.[48] Thus, it has been proposed that the *N*-glycosylation system in bacteria modifies flexible parts of proteins, and does not play a direct role in protein folding.[48]

The *Campylobacter* PglB protein is the first bacterial *N*-linked OT that has been experimentally verified, although many oligosaccharyltransferases have been identified with the sequencing of bacterial genomes.[39] Thus, the presence of an Stt3 homolog in the pgl locus suggests a glycosylation model analogous to *N*-linked glycan assembly in higher organisms. This finding may also support the proposals that both prokaryotic and eukaryotic glycosylation systems have evolved from a common ancestor.[46]

1.1.2.3 O-Glycosylation "Fla" in C. jejuni

The *Campylobacter* flagellin protein is one of the most extensively modified bacterial proteins characterized to date, with approximately 10% of the total mass of the glycoprotein consisting of *O*-linked carbohydrates.[39] Campylobacter flagellins were shown to bind to a sialic-acid-specific lectin, although it was later determined that the sugars on these flagellin are monosaccharide analogues of the related sugar, pseudaminic acid, which is a nine-carbon sugar and a member of the 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids.[45] Glycosylation is confined to 19 Ser and Thr residues in the central domain of the flagellin protein of *C. jejuni* (residues 81–176), in the region known to be on the surface of the flagellar filament.[45] As in eukaryotes, the sites of *O*-glycosylation are found at a distance from charged residues in hydrophobic patches of the target protein.[45]

Analysis of the gene cluster has identified 50 genes that may be involved in pilin *O*-glycosylation, however genetic analysis cannot find a GT in the cluster.[45] The same analysis identified multiple homologues of sialic acid biosynthetic genes in *C. jejuni* NCTC 11168 sequence and several of these genes, Cj1311 (neuA2), Cj1317 (neuB3), Cj1327 (neuB2), Cj1328 (neuC2), and Cj1331 (neuA3), reside in the flagellar glycosylation locus.[45] These genes have been determined to be involved in pseudaminic acid biosynthesis.[49] Other genes in the cluster remain hypothetical and are predicted to encode proteins involved in motility variation, or are proteins of unknown function.[24]

Bacterial *O*-glycosylation is believed to occur in the cytosol or inner membrane of the organism, with NDP-activated sugars as substrates.[24] Similar to eukaryotic *O*-glycosylation, it is proposed that *O*-glycans are assembled sequentially, and no lipid carriers are involved in this process.[24] Recent evidence also suggests that the *O*-antigen biosynthetic pathway may also be a source for pilin glycosylation precursors.[7] Examples

of shared enzymes have been found in multiple bacterial strains, including *E. col*i, *Haemophilus influenza*, *H. pylori*, *N. meningitidis*, *Aneurinibacillus thermoaerophilius*, and are proposed to occur in others.[24] Sharing enzymes between biosynthetic pathways such as those involved in pilin glycosylation and *O*-antigen biosynthesis may allow prokaryotes to avoid redundancy and maintain compact genomes.[24]

1.1.2.4 Bacterial O-Antigens

O-Antigens are important outer membrane components of Gram-negative bacteria, and are not glycopeptides.[50] The corresponding bacterial glycosylation pathway is briefly reviewed here not only due to the apparent functional similarities of the *O*-antigen to *O*-GalNAc derived glycans (mucins) and glycosaminoglycans (GAG) in higher eukaryotes but also due to its biosynthetic similarity to S-layer glycans in Gram-positive bacteria and archaea. The most extensively studied *O*-antigens are found in *E. coli* and *Salmonella enterica*, with 186 and 54 different antigens documented, respectively.[50] The diversity of structures precludes a succinct summary of biosynthetic pathways, if detailed descriptions of each structure were attempted. Thus, only a basic biosynthetic summary of these pathways and their roles in prokaryotic biology will be presented here.

1.1.2.5 O-Antigen Biosynthesis

O-Antigen biosynthesis can be divided into three steps, nucleotide sugar biosynthesis, sequential attachment of the activated sugars to the lipid carrier followed by membrane translocation, and final assembly of the *O*-antigen units to form the heteropolymer.[50] Common sugars such as, Glu, Gal, Rha, Man and their derivatives, and less common sugar moieties such as colitose, paratose, abequose, tyvelose, ascarylose

and others, are found in *O*-antigens.[50, 51] *O*-Chains of different organisms are very structurally diverse and are comprised of a variety of sugars.[52] Many biosynthetic genes for common sugars are found elsewhere in the bacterial genome, while biosynthetic genes for specialized sugars are located in the *O*-antigen gene cluster.[50] After sugar biosynthesis, sugar moieties are assembled sequentially to create the "O-unit" on undecaprenyl phosphate, which serves as the lipid carrier in the bacterial membrane.[50] The heteropolymer is generally flipped at this time by flippase Wzx, and polymerized by Wzy polymerase, which creates a 3-sugar main chain with 2-sugar side branch or a 2-sugar main chain with 3- sugar side branch.[50, 53] The polymer length is determined by Wzz, and the finished product is exported by Wzm and Wzt.[50] The finished product is anchored on Lipid A, and may be subsequently modified by further glycosylation or acetylation.[50, 54]

1.1.2.6 Roles of Prokaryotic Surface N- and O-Glycosylation

Early reports of bacterial glycosylation led to the investigation of other glycosylated cell surface appendages, such as pili and flagella.[7, 33, 55] The pili of the pathogenic bacteria *Neisseria gonorrhoeae*, *N. menigitidis*, and *P. aeruginosa* are composed mainly of the glycoprotein pilin.[56, 57] Pilin glycosylation genes were discovered in these bacteria and the encoded enzymes transfer rare sugars, such as a modified version of pseudaminic acid, a nine-carbon sugar that resembles sialic acid.[33, 49, 57] The pilins of gonococci and meningococci were found to be glycosylated at Ser63, but instead of GlcNAc, a novel saccharide, 2,4-diacetamido-2,4,6-trideoxyhexose, was identified as the *O*-linked sugar residue.[56, 58] Subsequent work in *Campylobacter jejuni* led to the identification of a similar trideoxyhexose, termed bacillosamine.[39]

Prokaryotic glycans appear to be more complex than initially supposed. Other than the cell envelope, glycosylated flagella are also important cell surface appendages of bacteria, as motility is a key factor in the adaptation of many bacterial pathogens.[6] These structures have been found in Gram-negative pathogens and methanogenic archaea, such as *M. voltae* and *H. halobium*.[7, 33, 59, 60] Interestingly, the *N*-linked glycans of the individual flagellins are similar but not identical to the Glc→Asn-linked glycan chains present in the halobacterial S-layer glycoprotein in *H. halobium*. Complex flagella assembled from two or more structural flagellin proteins have been found in a diverse range of Gram-negative bacteria, including *Caulobacter*, *Campylobacter*, *Aeromonas*, *Vibrio*, *Pseudomonas*, and *Helicobacter*.[57, 61] As in the pilin of *P. aeruginosa*, various derivatives of pseudaminic acid were identified in various bacterial organisms.[49] Mounting evidence also indicates glycosylation involvement in the flagellar assembly process.[57]

The function of bacterial *N*-glycosylation is not well documented. Loss of *N*-glycosylation does not render archaea inviable. Protein *N*-glycosylation in *C. jejuni* appears to play a role in protein complex formation or assembly in the inner membrane.[24] Nearly all *N*-glycosylated proteins in *C. jejuni* have a signal peptide, such that most of these proteins have an extracytoplasmic location.[24] Possible functions of these proteins include protection from proteolytic cleavage, protein stability, or perhaps serving as cellular sorting signals.[24, 62]

O-Antigens are found in both pathogenic and non-pathogenic bacteria. Certain types of *O*-antigens are major virulence factors in infectious bacteria, exhibiting pyrogenicity and lethal toxicity in higher organisms. They are also important for bacterial survival.[52] For example, *E. coli* strains that have acidic sugars in their *O*-antigens often cause dysentery-like diseases.[63] GalNAcA(N), a rare acidic sugar, is also present in the *P.*

aeruginosa O6 *O*-antigen.[63] *O*-Antigens such as O157, O111 and O26 associated with EHEC (enterohaemorrhagic *E. coli*) are of particular significance in human diseases.[64] First isolated from infantile diarrhea, O26 and O121 strains were recently implicated in serious enteric disorders, such as hemolytic uremic syndrome.[63, 64] *E. coli* O26 has a low infectious dose, and is a major concern in food-borne illness, as cattle are the main bacterial reservoir.[64] *E. coli* O121 strains have been isolated from multiple clinical specimens and have been implicated in a waterborne outbreak of infection.[63] *Shigella dysenteria* O7 antigen was recently found to have the same structure as the *E. coli* O121 antigen, thus suggesting that common *O*-antigen structures may result in common disease states, even though the bacterial sources of the antigens may differ.[63]

In summary, the role of bacterial protein glycosylation is as varied as glycosylation in eukaryotes, and many subjects of research remain to be investigated.

1.1.3 BACTERIAL NATURAL PRODUCTS

Turning now to a completely different form of bacterial glycosylation, some bacteria produce and secrete bioactive secondary metabolites. These natural products take many forms, and may be antibiotics, virulence factors, or toxins.[65] Bioactive natural products may also be produced by marine organisms or plants, although here we confine our discussion to bacterial natural products, and specifically the role of glycosylation in the activity of these compounds.

The majority of natural products produced by bacteria follow several routes of biosynthesis including isoprene assembly, non-ribosomal peptide (NRPS) biosynthesis or polyketide (PKS) assembly.[65] Excellent reviews on these biosyntheses are readily available and will not be discussed in this work.[66, 67] Many of these secondary metabolites

are decorated with various deoxysugars that provide an additional level of complexity to the intricate natural product scaffolds.[65] The addition of deoxysugars through glycosylation can have a dramatic impact on the biological activity, bioavailability and pharmacology of these compounds. [65] Glycosylated natural products include macrolides such as erythromycin, peptides such as vancomycin, and aminoglycosides such as gentamicin, and spirotetronates such as kijanmicin.[67, 68] Glycosylation is also important for the activity of natural product anticancer agents such as doxorubicin, antiparasitic agents such as avermectin, and antifungal agents such as blasticidin-S.[65]

Carbohydrate moieties of bioactive compounds have been implicated in the control of drug pharmacokinetics including absorption, distribution and metabolism.[69] Evidence also suggests that sugar components are often involved in interaction between drug and cellular target and their presence is important for the bioactivity of many natural products.[69, 70] A growing number of natural products produced by bacteria consist of oligosaccharide chains.[71] These chains are composed of deoxysugars that are further modified by methylation, acetylation, and amination.[71]

Bacterial natural products require NDP-activated sugars for glycosylation, and most often these sugars are transferred by a dedicated glycosyltransferase to a nucleophilic receptor (hydroxyl or amino group) on the aglycone scaffold.[65] The resulting *N*- and *O*-linkages are quite common in bacterial natural products and *O*-glycosylation in particular has been extensively studied.[72] *C*-glycosylation is rarely found in bacterial natural products, and is of particular interest from the pharmaceutical viewpoint since these bonds are stable towards glycosidase degradation.[71] This rare type of glycosylation has been identified in the antibiotics medermycin, granaticin, and the anticancer agent urdamycin.[65] *C*-glycosylation is more commonly found in aryl-*C*-glycosides, which are biologically important natural products including pluramycins, anguicylines,

benzoisochromanequinones (BIQ) and salmochelin siderophores.[73, 74] In pluramycins, anguicylines and BIQs, *C*-glycosidic linkages are located ortho to a hydroxyl group, while *C*-glucosylation on salmochelins occurs at 2,3-dihydroxybenzoyl rings at position 5, para to a phenol hydroxyl group.[73, 75]

1.1.3.1 Natural Product Glycosyltransferases: Classification, Folds and Mechanisms

Glycosyltransferases catalyze the transfer of a sugar moiety from an activated donor to an acceptor molecule, which is termed the aglycone in natural products.[71] Glycosyltransferases are classified using a system based on substrate specificity, type of reaction, and amino acid sequence similarity of GTs.[71] However, despite their apparent diversity in functions, they demonstrate great structural similarity.[71, 76] Actinomycetes, which produce the majority of glycosylated bacterial secondary metabolites have glycosyltransferases that, for the most part, belong to the GT-1 family (CAZy classification).[71] Glycosyltransferases from strains producing polyketide antibiotics are more similar to each other than enzymes from strains producing glycopeptides or polyene antibiotics. In particular, glycosyltransferases from the macrolide pathways are shown to be quite closely related.[71] Phylogenetic analysis of glycosyltransferases also demonstrates relationships between enzymes that act on the same region of similar aglycones, regardless of the identity of the NDP-sugar substrate.[71]

Two types of folds have been observed via X-ray analysis of glycosyltransferases, GT-A and GT-B.[71] GT-A has two dissimilar domains. The *N*-terminal domain is involved in nucleotide sugar binding, comprising several β-strands that are each flanked by α-helices as in an α/β/α sandwich topology which resembles a Rossmann fold. The *C*-terminal domain, which contains the acceptor-binding site, consists largely of mixed β-

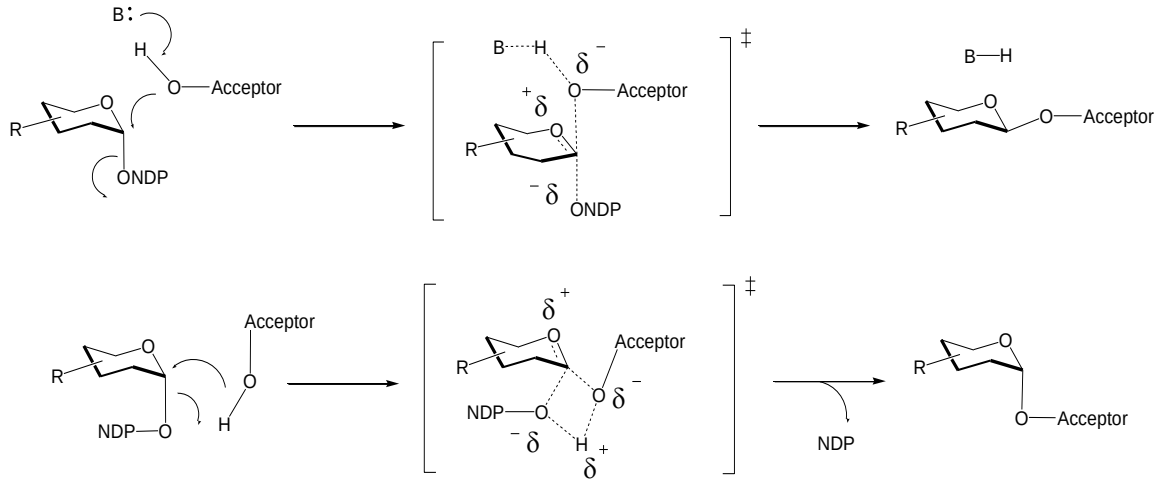
sheets.[77, 78] The NDP-binding region of GT-A enzymes generally contains a DXD motif that is involved in metal binding and anchoring the diphosphate moiety of the NDP-sugar.[78] Members of the GT-B superfamily are also bi-lobal enzymes that have folds consisting of two β/α/β Rossmann-like domains; with the active site lying in the cleft between the domains.[79] In contrast to GT-A family enzymes, in GT-B family enzymes, the *N*-terminal domain provides the acceptor (aglycone) binding site, while the *C*-terminal of the enzyme is associated with either the NDP-sugar or the donor substrate. [79] GT-B family enzymes are more disparate than the GT-A family proteins, sharing less than 10% sequence identity, although structures that have been solved demonstrate similarity.[78] It should be noted here that the fold or superfamily of the enzyme does not dictate the mechanism of the enzyme or the resulting stereochemical outcome of the product. [80] Indeed, non-glycosyltransferase enzymes are also known to adopt the GT-B fold, an example of which is UDP-GlcNAc 2-epimerase.[81]

The majority of natural product glycosyltransferases are members of the GT-B family.[78] The members of the GT-B superfamily seem to display a greater degree of diversity in the glycosyltransferase catalytic mechanisms between and among family members, as compared to what has been observed for inverting GT-A enzymes.[79] The greater mechanistic diversity is perhaps a reflection of the greater diversity of chemistries catalyzed by this superfamily of enzymes, and perhaps also by the physical separation of the two domains. Thus, the mechanisms of these enzymes cannot be definitively predicted, although a general mechanism can be described.

Briefly, the S_N2 inverting mechanism involves the direct attack of the anomeric center, displacing the NDP activating group. In this mechanism, the active site nucleophile is proposed to stabilize and position the aglycone in the active site for the direct attack, while the NDP group is again displaced with the assistance of a metal ion.[79]

As the NDP group is displaced, an active site base abstracts a proton from the aglycone hydroxyl.[79] The resulting hydroxyl ion can attack the opposing face of the sugar, finally transferring the sugar to the aglycone and inverting the stereochemistry of the bond. Thus, the nucleophilic acceptor attacks the anomeric carbon from the opposite side of the sugar ring as the NDP leaving group in a synchronous, concerted manner with highly dissociative oxocarbenium-like character. [82]

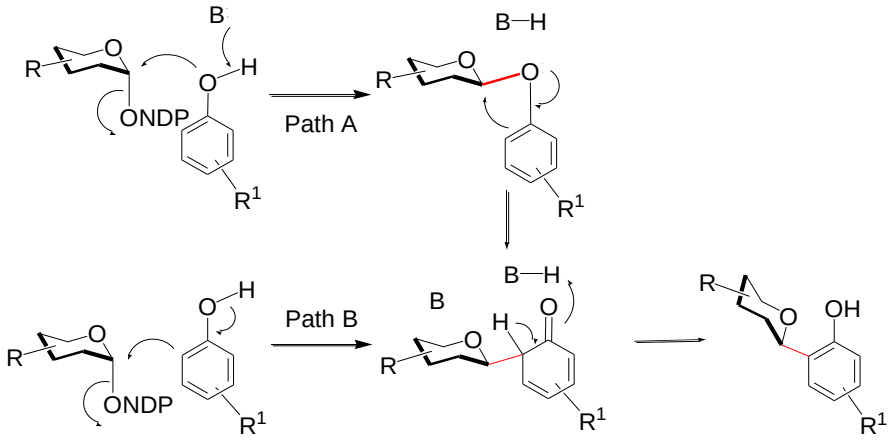
1.5 Mechanisms of Inverting and Retaining Glycosyltransferases



Meanwhile, the S_N1 retaining mechanism is proposed to pass through a short-lived oxocarbenium transition state, which is facilitated by the lone pair electrons from the endocyclic oxygen. As the NDP group is displaced with the assistance of a metal ion, the terminal phosphate abstracts a proton from the aglycone hydroxyl.[79] The NDP group displacement occurs without the stabilization of an active site nucleophile, a transient

oxocarbenium ion forms. The formation of the oxocarbenium ion enables the direct attack of the anomeric carbon by the aglycone hydroxyl ion, on the same face as the leaving group, thus retaining the stereochemistry of the original NDP group. [83] These mechanisms are shown in figure 1.2.

1.6 Proposed *C*-Glycosylation Mechanisms



The mechanism of *C*-glycosylation is not as well studied as *O*-glycosylation, although there are two current mechanistic proposals. Due to the similarity of O- and C-glycosyltransferases, a Fries-type rearrangement was proposed.[73] The mechanistic rearrangement would begin with *O*-glycosylation of a hydroxylated aglycone, after which an intramolecular rearrangement would transfer the O-linkage to a C-linkage ortho to the original hydroxyl group of the aromatic ring.[73] The second proposed mechanism is a direct alkylation, via a Friedel-Crafts type mechanism, requiring an activated glycoside donor and an aryl acceptor that is activated through its inherently electron-rich nature.[73] Both mechanisms are shown in Figure 1.3.

Synthetic preparation of aryl-*C*-glycosides by Friedel-Crafts type reaction resemble natural aryl-*C*-glycosylation with ortho- and/or para-substituted products predominating.[73] For example, in salmochelin, selective glucosylation occurs exclusively at position 5, which is para to the phenol hydroxyl at position 2 of the 2,3-dihydroxybenzoyl group.[75] In many other *C*-glycosylated natural products the addition of sugars is ortho to the phenol hydroxyl, again mirroring the Friedel-Crafts type of coupling.[73]

The simplicity of a direct nucleophilic attack of an activated glycoside by an electron-rich aromatic ring makes this an attractive and feasible mechanism. Supporting this model, the UrdGT2 glycosyltransferase form the urdamycin biosynthetic pathway was found to catalyze both *O* and *C*-glycosylation in a heterologous system using *S. fradiae*.[84] This study was the first to demonstrate that a natural product GT could synthesize both *C*- and *O*-glycosidic linkages.[84] These results support the direct *C*-glycosylation mechanism, because in this case, initial *O*-glycosylation of the aglycone followed by O-C rearrangement is not favored.[78] Moreover, the recently solved X-ray crystal structure of UrdGT2 revealed that the anomeric carbon of the NDP-sugar substrate binds in close proximity to C-9 of the aglycone substrate, and is properly positioned for a direct addition to the aromatic ring.[85] Asp137 was proposed to be the base responsible for deprotonation of the aglycone phenolate group via a tightly bound active site water molecule.[78]

1.1.4 PROKARYOTIC GLYCOSYLATION SUMMARY

As shown above, prokaryotic glycosylation is incredibly varied and includes the modification of proteins, lipids and small molecules. These compounds and

macromolecules have diverse biological activities, and may be essential for prokaryotic viability. These glycosylation events may be superfluous, however, most provide at least some benefit to the organism. Glycosylation includes *N*- and *O*-linkages on proteins, lipids and small molecules. Rarely, *C*-linkages may occur on small molecules. It is interesting to compare these diverse glycosylation events to eukaryotic systems, because this comparison can provide information regarding both prokaryotic and eukaryotic systems.

1.2 ROLE OF EUKARYOTIC GLYCOSYLATION

The biological roles of glycosylation in multi-cellular eukaryotes are complex, varied and difficult to define in terms of absolute functions. In fact, glycosylation is the most complex type of posttranslational modification in eukaryotes.[86] Although, the roles that glycosylation may play in the human body have not yet been fully investigated, many theories have been posited regarding the roles of each type of glycosylation.[87] However, for every rule there are exceptions that can be found.[87] It is generally believed that carbohydrates in multicellular organisms have broad-reaching roles, ranging from intracellular sorting, protein quality control, cellular communication, cellular adhesion, immune response, to host-pathogen interactions.[87, 88] Perhaps the most inclusive statement regarding the biological role of oligosaccharides is that they either mediate 'specific recognition' events or they provide 'modulation' of biological processes.[87] It should be noted that 0.5 to 1% of the transcribed human genome is devoted to the production of proteins involved in the synthesis, degradation, and function of glycoconjugates, therefore these sugar moieties may play a more significant role in the biology of multicellular organisms than was previously thought.[89]

Across eukaryotic species, there are eight amino acids and 13 monosaccharides that participate in protein glycosylation, giving rise to 31 different types of glycosidic linkages.[90] When taking into account anomeric configurations of the sugars, this number rises to 37.[90] However, the majority of eukaryotic protein glycosylation can be divided into two groups, *N*-linked and *O*-linked, comprising five and 27 of 37 types of linkages respectively across species.[88, 90, 91] Among the 31 types of potential glycosidic linkages in eukaryotes, the majority of *N*-glycans are of one type (β-Asn), while twelve types of *O*-glycans have been identified.[91] These two types of protein glycosylation are associated with many congenital diseases (congenital defects in glycosylation) in humans, thus they are the most well-characterized.[87, 89, 91, 92]

These two types of glycosylation, which are conserved from yeast to humans, share many similarities.[88, 91] First, in the case of many organisms, glycosylation is absolutely required for normal development.[87, 89] Complete loss of *N*-glycosylation is lethal in both yeast and mammals, while partial loss results in severe multi-systemic defects and abnormalities leading to severe or lethal phenotypes.[87, 89] It appears that glycosylation plays a major role in embryonic and postembryonic development, thus it has been difficult to study the effects of glycosylation-defective intact multi-cellular organisms.[87, 91] Work by Robbins and co-workers in identification of genes involved in asparagine-linked glycosylation (*alg* genes) allowed studies on *N*-linked glycosylation in yeast.[46] These studies demonstrated that if such *alg* mutations were combined with a conditional mutation affecting oligosaccharyltransferase activity in the same cell, a lethal phenotype resulted. [93]

Second, both *N*-linked and *O*-linked glycosylation begin in the endoplasmic reticulum (ER) of the cell, and progress through the Golgi apparatus.[90, 91] The reactions of *N*-glycosylation in the ER hardly differ between yeast and animals, however

subsequent reactions, taking place in the Golgi cisternae, differ considerably.[91] The biosynthesis of sugar chains linked through *O*-mannoside also proceeds in mammals in the same way as in yeast, as was recently shown.[94-96] Mucin type *O*-glycosylation, thought to be limited to eukaryotes, begins in the cis-Golgi apparatus across species.[90, 91] Current research suggests that many forms of glycosylation are conserved in manner, location and pattern.

1.3 SELECTED TYPES OF PROTEIN GLYCOSYLATION, *N*-GLYCOSYLATION AND MUCIN TYPE *O*-GLYCOSYLATION (*O*-GALNAC DERIVED GLYCOSYLATION)

1.3.1 *N*-GLYCOSYLATION

Asparagine linked glycosylation was first discovered in the 1960s, and the oligosaccharide structure was elucidated by the work of Kornfeld and coworkers. An excellent review of this work can was published in Annu. Rev. Biochem. In 1985.[97] *N*-linked glycosylation is characterized by β-glycosylamide linkage to asparagine.[98] As mentioned previously, in all eukaryotes, the *N*-glycosylation process is conserved through the endoplasmic reticulum (ER). This conservation includes the construction of the core *N*-glycan structure on the substrate lipid, which is then transferred *en bloc* to the protein acceptor by a complex of enzymes known as the OST, or oligosaccharyltransferase. [97]

In prokaryotic *N*-glycosylation systems, polyprenols continue to serve as lipid carriers and the oligosaccharide is also constructed on the lipid, followed by transfer to the protein.[46, 98] A comparison of the *N*-linked glycosylation machinery involved in eukaryotic systems and the model prokaryotic system found in the Gram-negative bacteria *C. jejuni*, similarities become readily apparent.[98] Prokaryotes need undecaprenyl

as a lipid substrate while dolichol is utilized in higher organisms.[46] Several proteins involved in the biosynthesis of the lipid-linked oligosaccharide (LLO) show sequence similarity to eukaryotic and prokaryotic proteins.[46] This finding is not unexpected when the similarity between lipid-linked oligosaccharides (LLO) in eukaryotes and lipopolysaccharide (LPS) biosynthesis in bacteria is considered. [99]

For example, *N*-acetylglucosaminyl phosphate transferase (GPT), the protein catalyzing the first step in LLO biosynthesis, displays a signal sequence similar to bacterial UDP-N-acetylmuramoyl-pentapeptide transferases involved in murein biosynthesis, and UDP-GlcNAc:undecaprenol-P GlcNAc-1-P transferases involved in lipopolysaccharide biosynthesis.[46] Both the ALG1 and the ALG2 proteins, the key proteins in eukaryotic *N*-linked-glycosylation have sequence similarity to bacterial glycosyltransferases.[46] Alg1p and Alg2p share a common sequence motif with these bacterial glycosyltransferases. This motif is also found in the PIG-A protein, a GlcNAc transferase required for the assembly of the GPI-anchor precursor, or the sucrose synthase of plants.[100, 101] The existence of a conserved sequence motif in such diverse glycosyltransferases which all utilize nucleotide-activated sugars suggests a common evolutionary origin of these enzymes.[46]

There are some key differences between eukaryotic and prokaryotic systems. In prokaryotic systems, these glycosylation events must occur on the bacterial membrane. In eukaryotic systems the membrane of the endoplasmic reticulum is the location of these glycosylation events, and events occur on both sides (cytosolic and luminal) of the membrane.[46, 97] Thus, in eukaryotes, a flipping of the lipid-linked oligosaccharide across the membrane is a necessary step.[46] In the dolichol pathway assembly of the oligosaccharide continues in the lumen of the ER after the flipping reaction, whereas in lipopolysaccharide biosynthesis the oligosaccharide is transferred from the carrier

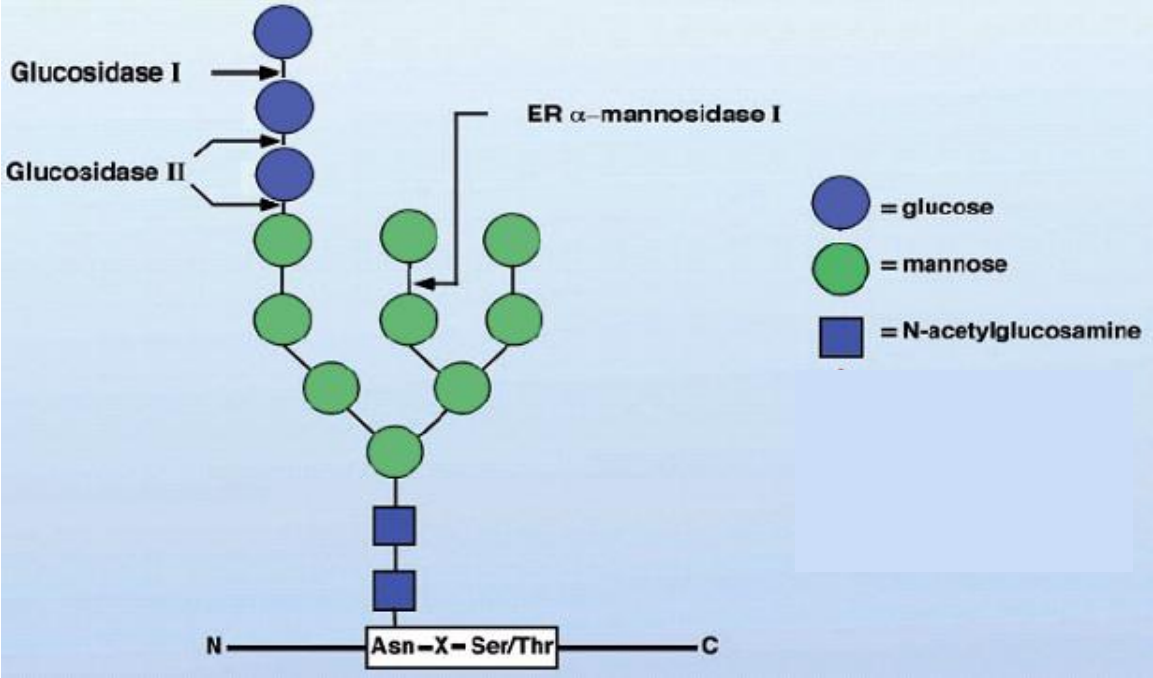
undecaprenol to lipid A in the periplasmic space.[102] The glycans formed in each process are different, prokaryotes forming a much shorter non-branched chain, while a longer branched chain is formed in eukaryotic systems.[46, 97] Some of the bacterial glycosyltransferases that are similar to eukaryotic proteins involved in *N*-glycosylation use different nucleotide-activated sugars and are involved in the biosynthesis of lipopolysaccharides or exopolysaccharides, instead of protein glycosylation.[46] These glycosyltransferases required for the assembly of lipopolysaccharides or exopolysaccharides in both Gram-negative and -positive bacteria may represent a distinct family of glycosyltransferases.[46]

A sequence search for Dol-P-man or Dol-P-glc dependent glycosyltransferases Alg3, Alg9, Alg6, Alg8, and Alg10, which are involved in luminal biosynthetic reactions, found no similarity to any bacterial proteins.[46] These results led to speculation that the first part of the eukaryotic dolichol-linked oligosaccharide biosynthesis, which takes place at the cytoplasmic side of the ER membrane, shares a common origin with bacterial lipopolysaccharide biosynthesis.[46] *N*-linked glycosylation also occurs in archaea, and has been shown to follow a similar pathway to that of eukaryotic cells.[5, 60, 103] Eubacterial glycosyltransferases have been shown to share a tripartite sequence motif found in Dol-PP-glc and Dol-P-man synthases.[46] Intriguingly, proteins similar in predicted structure (membrane proteins with C-terminal hydrophilic domain) and in primary sequence to eukaryotic STT3 protein have been detected in all three archaeal genomes sequenced thus far.[104] Taken together, these findings suggest that eukaryotic *N*-linked protein glycosylation may originate from an archaebacterial process, that in turn is derived from a similar process in archaea. [46]

1.3.1.1 N-glycan Overview

The *N*-oligosaccharide structure was elucidated by the work of Kornfeld and coworkers, and subsequent research has established the biosynthetic pathway of *N*-glycan formation and attachment.

1.7 The Core *N*-Glycan Structure



As previously mentioned, the glycan is assembled in the ER on the lipid carrier dolichol phosphate.[97] Snider and co-workers used ConA to define the membrane orientation of the lipid-linked oligosaccharide intermediates and the enzymes involved in their synthesis.[105, 106] Their data indicate that Man3-5GlcNAc2 species face the cytosol whereas the Man6-9GlcNAc2 and Glc1-3Man9GlcNAc2 species face the lumen of the ER.[97, 105, 106] These results suggest that the Man5GlcNAc2 species is assembled on the

cytosolic face of the ER leaflet and then translocated to the luminal side, where the assembly continues.[97, 105, 106]

1.3.1.2 N-Glycosylation: The Cytosolic Reactions

Thus, the first steps of LLO biosynthesis occur on the outer membrane, or cytosolic face of the ER. Here, sugars are added in a stepwise fashion with the first seven sugar moieties (2 *N*-acetylglucosamine (GlcNAc), and 5 mannose (Man)) derived from the nucleotide activated sugars UDP-GlcNAc and GDP-Man.[97] Biosynthesis of the LLO begins on the cytoplasmic side of the ER with the addition of GlcNAc to Dol-P by Alg7. The heterodimeric complex Alg13/14 catalyzes the second GlcNAc addition to the LLO.[107, 108] In the third step, the *alg1* gene product transfers GDP-Man to Dol-PP-$GlcNAc_2Man_1$.[46] The fourth cytoplasmic glycosylation step is catalyzed by Alg2, which adds two mannose moieties to the growing chain to generate Dol-PP-$GlcNAc_2$-Man_3 [109] The fifth and final cytoplasmic step is catalyzed by Alg11, which also adds two mannose moieties to the growing chain to generate Dol-PP-$GlcNAc_2$-Man_5.[109] The glycan chain is flipped such that the glycan is now within the lumen of the ER where subsequent additions occur.

To obtain the NDP-sugar substrates, GlcNAc and Man glucose-6-phosphate (Glc-6-P) is converted to fructose-6-phosphate (Frc-6-P), catalyzed by phospoglucoisomerase.[110] Fructose-6-phosphate is subsequently converted to glucosamine, then acetylated to form *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P).[46, 111] GlcNAc-6-P is converted to *N*-acetylglucosamine-1-phosphate (GlcNAc-1-P) by the enzyme *N*-acetylglucosamine-phosphomutase and finally to UDP-*N*-Acetylglucosamine (UDP-GlcNAc).[46, 112] The GDP-mannose biosynthetic pathway is well-characterized.[46]

To generate GDP-mannose, fructose-6-phosphate is isomerized to mannose-6-phosphate (Man-6-P) by phosphomannoisomerase (PMI40).[113] Mannose-6-phosphate is then activated by transfer to GDP by GDP-Man pyrophosphorylase to form GDP-mannose (GDP-Man).[114, 115]

1.3.1.3 N-Glycosylation Substrates: Lipid-Linked Sugars and the Role of Dolichol-Phosphate

The next seven sugars added to the growing glycan chain are derived from lipid intermediates in the lumen, dolichol-phosphate-mannose (Dol-P-Man) and dolichol-phosphate-glucose (Dol-P-Glc).[97] Dolichol-phosphate (Dol-P) acts as the carrier in the assembly of pyrophosphate-linked oligosaccharides and is the acceptor in the synthesis of the sugar donors Dol-P-Man and Dol-P-Glc from GDP-Man and UDP-Glc, respectively.[46] Synthesis of Dol-P-linked oligos requires UDP-GlcNAc, UDP-Glc, and GDP-Man as direct donors for LLO synthesis in the cytosol, or as indirect donors to generate Dol-P sugars that serve as substrates after LLO flipping into the lumen of the ER.[46]

Dol-P is the substrate of a number of enzymes in the *N*-glycosylation pathway and the availability of this lipid is one of the rate-limiting factors in the synthesis of LLOs in higher eukaryotic cells.[116-119] Dol-P is generated by *de novo* synthesis but it is also a product of the Dol-P-Man and Dol-P-Glc dependent glycosyltransferase reactions.[46] The dolichol cycle of recycling from the lumen of the ER to the cytosol and how it occurs remains to be understood.[46] Recycling would explain the regulation of a pathway which takes place at both the cytoplasmic and luminal sides of the ER.[46] Flux through the pathway would then be regulated by the availability of dolichol or Dol-P.[46] Lipid linkage may be the method of choice by which cells can "overcome the barrier of diffusion

control and to make the multimolecular reaction processes at low concentrations more efficient". [120] The lipid linkages may also facilitate the very rapid rate of LLO biosynthesis ensuring the availability of fully assembled LLO that is required for the efficient *N*-linked glycosylation of translocating proteins.[46]

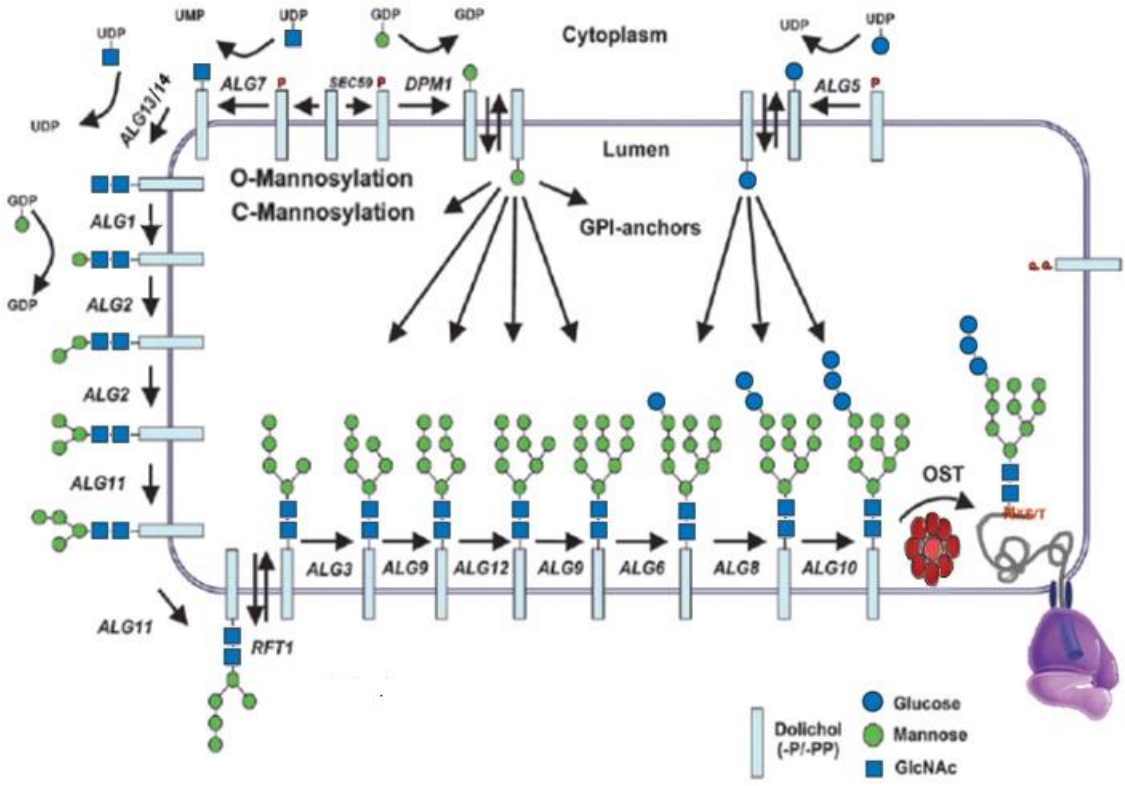
Other than Dol-P, the requisite donor sugars GlcNAc, Man and Glc must also be available for *N*-glycosylation to proceed. To generate UDP-glucose, glucose-6-phosphate is converted to glucose-1-phosphate by phosphoglucomutase (*PGM1*, *PGM2*), in preparation for nucleotidyl activation.[121, 122] Glucose-1-P is activated to UDP-glucose by UDP-Glc pyrophosphorylase (UGPase). The gene encoding UGPase, *UGP1*, is an essential gene. The generation of the remaining two sugars, GlcNAc and Man, was discussed in the previous section. Once the NDP-sugars are biosynthesized, some of these sugars are transferred to the dolichol lipid in the membrane of the ER. The *alg5* gene encodes Dol-P-Glc synthetase, and it is proposed that the Dol-P-Glc is generated on the cytoplasmic side of the ER and later flipped into the lumen.[123] GDP-Man is transferred to dolichol to form Dol-P-Man by *DPM1*, which is an essential gene for cellular viability. Dol-P-Man is a substrate used in many cellular glycosylation processes including *N*-linked glycosylation, *O*-linked glycosylation, GPI anchors and, as will be discussed later, *C*-linked glycosylation.[46, 95, 124-126] Notably, steps in *N*-glycosylation requiring Dol-P-Man are not essential, as *alg3* mutants accumulate the Alg3 substrate, $Man_5GlcNAc_2$-PP-Dol, and this glycan can be transferred to protein in cells devoid of any Dol-P-Man.[127]

1.3.1.4 N-Glycosylation: The Lumenal Reactions

Using these Dol-P substrates, $Man_5GlcNAc_2$-PP-Dol is elongated in the lumen of the ER. A total of 4 mannosyltransferases and 3 glucosyltransferases are involved in the

completion of the LLO core oligosaccharide, which are responsible for the additions of 4 mannoses followed by 3 glucoses in a stepwise fashion at the luminal side of the ER membrane.[46, 128, 129]

1.8 The *N*-Glycosylation Pathway



The first step in luminal biosynthesis is catalyzed by the Alg3 enzyme, which initiates the Dol-P-monosaccharide-dependent glycosyltransferase reactions at the luminal side of the ER (α-1,3-mannosyltransferase).[46, 127] The subsequent step is catalyzed by Alg9, an α-1,2 mannosyltransferase that adds a mannose residue to the α-1,3 Man.[97, 128] The subsequent addition of and α-1,6 linked mannose is postulated to be

catalyzed by the Alg12 enzyme, and Alg6 is believed to catalyze the transfer of the first glucose to the oligosaccharide (α-1,3).[46, 130] Alg8 protein adds the second glucose, and Alg10 adds the terminal glucose to the oligosaccharide in α-1,3 and α-1,2 linkages respectively, finally yielding $Glc_3Man_9GlcNAc_2$.[46] These glycosyltransferases that utilize Dol-P-mannose or Dol-P-glucose are highly hydrophobic, basic proteins, with calculated molecular mass between 60-75 kDa and are assumed to contain multiple transmembrane domains with only small hydrophilic loops.[46]

The LLO elongation is a highly ordered assembly in which individual sugar residues of the protein bound core oligosaccharide fulfill specific functions in the processing of glycoproteins in the ER.[131] These roles may explain the high level of conservation and the precision of the assembly pathways of the LLO in eukaryotic cells.[46] All *alg* genes have homologs in higher organisms (around 50% identity), and for each step in biosynthesis a single glycosyltransferase exists, with the exception of *alg2* and *alg11,* which each catalyze two steps of the LLO elongation.[93, 109, 123, 127, 130, 132] These late glycosyltransferases in lumen use Dol-P-saccharides and are non-essential for viability as mutations in these enzymes result in underglycosylation of secretory proteins in vivo.[46] Underglycosylation appears to be due to a decreased affinity of oligosaccharyltransferase (OST) towards truncated oligosaccharides. These glycans are still transferred, just with reduced efficiency.[46]

1.3.1.5 N-glycosylation: Protein Glycosyltransferase Oligosaccharyltransferase Complex

The oligosaccharyl transferase (OT) complex consists of nine subunits, five of which are essential gene products.[133] The OT catalyzes the *N*-glycosylation of most secretory and membrane bound proteins.[133] The OT complex transfers the Dol-PP-

$Glc_3Man_9GlcNAc_2$ to the asparagine residue in the protein consensus sequence of NXS/T (where X is any amino acid) on a growing polypeptide chain.[62] Statistical analysis indicates that only 65% of all available *N*-glycosylation sites are utilized.[134]

Three subcomplexes have been proposed, comprised of: Ost1p-Ost5p, Ost2p-Swp1p-Wbp1p, and Stt3p-Ost4p-Ost3p, although studies do not completely support this proposal.[135] Maximal rate of N-glycosylation requires presence of all subunits.[135] Ost1p, Ost2p, Wbp1p, Swp1p, and Stt3p are proposed to carry out critical functions.[135] Subunits encoded by non-essential genes (Ost3p, Ost6p, Ost4p, Ost5p) may perform important accessory functions.[135] Each protein component in this complex will be examined in turn.

1.3.1.6 Ost1p/Ribophorin I

Ribophorin I was found to crosslink with all other OT subunits and Alg 1, an enzyme involved in Dol-PP-oligosaccharide biosynthesis and to ribosomes.[133, 136, 137] Ribophorin I was also found to crosslink to a subset of membrane protein, irrespective of their glycosylation status after integration into the Sec61 translocon.[138] Ost1p was also able to crosslink to glycosylatable substrate-based peptide photoaffinity probe. Thus it was proposed that Ost1p bears the peptide-binding site of the OT complex. However, it was shown later that this is not the case.[139] Instead, it is now postulated that the luminal domain of ribophorin I may be involved in funneling the newly synthesized peptides into the active site of Stt3p for the *N*-glycosylation reaction.[133, 137] Ost1p has also been reported to be *N*-glycosylated at four consensus sites.[133]

1.3.1.7 Ost5p

Ost5p is a 9.5 kDa membrane protein, present in highly purified OST preparations.[140] This protein is not essential for growth, but its depletion results in a minor defect in OT activity. Ost5p is not known to interact directly with other OT components, most likely Ost1p and/or Stt3p.[140] No function is currently proposed for this protein.[137]

1.3.1.8 Ost2p

The Ost2p protein is predicted to possess three transmembrane segments with the *N*-terminus in cytosol and the *C*-terminus in the lumen.[141] The protein sequence is highly conserved, especially in transmembrane regions, suggesting that functional properties reside in these segments. Ost2p has been found to interact strongly with Wbp1p, which is implicated in Dol-PP-oligo recognition, and thus may aid Wbp1p in recognition of the Dol-pp-oligo.[136, 141]

1.3.1.9 Swp1p

The Swp1p protein is a product of an essential gene, but the function of this protein is yet unclear.[137] The protein has three transmembrane segments, with the *N*-terminus in the lumen and *C*-terminus in the cytosol.[135] Swp1p was found to interact and crosslink to components of the OST complex as well as the protein translocation complex, specifically Wbp1p, Sss1p and Stt3p.[136, 142, 143] These interactions led to the proposal that Swp1p may aid Ost1p in funneling nascent polypeptide chains to the active site of the OT complex, as the luminal domain of the protein interacts with Ost1p.[144]

1.3.1.10 Wbp1p/OST48

The Wbp1p protein is proposed to be responsible for the recognition of the substrate dolichol-PP-oligosaccharide.[137] Investigations using mammalian systems have shown that a chemically reactive hexapeptide could be covalently linked to either ribophorin I (homolog of Ost1p) or OST48.[145] Chemical modification of cysteine residues suggested that Wbp1p may be involved in recognition of the Dol-P-oligosaccharide.[146] Wbp1p has a GIFT domain, which in other proteins is known to bind oligosaccharides.[137] The Lennarz group has shown that the sequence of the luminal half of the protein is important for the function of Wbp1p.[144] They have also reported that the luminal domain may possess a divalent metal binding site.[144] It should also be noted that two *N*-glycan sites on Wbp1p are occupied.[133]

1.3.1.11 Ost4p

The Ost4p protein is a minimembrane protein of only 36 amino acids, and has been found to be encoded by a non-essential gene.[147] The null strain exhibits a temperature sensitive phenotype that results in a severe underglycosylation defect.[147] The protein functions in the binding of Ost3p and Stt3p together in a subcomplex of OT via interactions with amino acid residues near the cytosolic leaflet of the ER membrane.[148, 149] Ost4p has also been proposed to be responsible for recruiting Ost3p or Ost6p into the OT complex.[150]

1.3.1.12 Ost3p and Ost6p

These proteins are thought to have redundant functions, due to their sequence similarity and similar hydropathy plots.[137, 143] Disruption of either of the genes results in a

minor glycosylation defect, while a double mutant yields a severe underglycosylation phenotype that affects both membrane and soluble glycoproteins.[143] This redundancy is supported by the finding that two OT isoforms exist in the ER membrane that contain all OT subunits except Ost3p and Ost6p.[151] The two isoforms differ in that they contain either Ost3p or Ost6p.[151] Both proteins have lumenal domains predicted to have a thioredoxin-like fold.[152] The Sbh1p protein, the β subunit of the Ssh1 translocon, exhibits a specific interaction between Ost3p, not 6p, while the Sbh2p interacts with Ost6p, not 3p.[137, 151] All other OT subunits have an identical interaction pattern with the Sbh1p and Sbh2p proteins, which suggests that the two complex isoforms are generated by virtue of association with Ost3p or Ost6p and specifically associate with the similar translocon complexes.[137]

1.3.1.13 Stt3p

Three lines of evidence clearly indicate that this protein bears the active site of the OT region.[133] Mutagenesis studies using a labeled acceptor substrate indicate that the Stt3p protein bears the peptide recognition site and/or the catalytic site for oligosaccharyl transfer.[47, 137, 139, 153] The protein pglB, a known homolog of Stt3 in *C. jejuni* has been shown to be essential for glycosylation in this species.[41, 154] As discussed previously, no other OT subunits are present in *C. jejuni* that could account for the *N*-glycosylation activity.[44] In the mammalian system, Stt3p was the only protein that could be crosslinked to the glycosylation consensus sequence of the translocating nascent chain.[155] Stt3p has also been shown to be involved in cell wall biosynthesis in yeast β-1,6-glycan, and to interact with protein kinase cascade components via its *N*-terminal domain, which is oriented towards the cytosol.[133]

Stt3p was found to be *N*-glycosylated at two different sites, and abolishing this glycosylation at N_{535} or N_{539} resulted in lethal phenotype in yeast, although the loss of glycosylation does not prevent Stt3p incorporation into the OT complex.[133] These *N*-glycosylation sites are highly conserved across species, although only N_{539} was found to be *N*-glycosylated in the wild type enzyme.[133] Taken together, these data suggest that *N*-glycosylation is required for the function of Stt3p, the enzyme responsible for *N*-glycosylation.[133] The function of Stt3p *N*-glycosylation was proposed based on the secondary structure prediction, as these glycosylation sites (N_{535} and N_{539}) are located on a flexible loop that may be possible to interact with the catalytic site.[133] *N*-glycan sites located less than 12-14 residues from a transmembrane segment of a protein are not modified *in vitro*, which suggests that the active site of the OT is located 30-40 Å from the luminal face of the ER.[133]

Lennarz and co-workers have also proposed that the N_{539}NT glycans may be close to the catalytic site W_{516}WDYG via interaction between the glycans and the aromatic amino acids W_{516}W, and that this interaction may be important for the function of Stt3p.[133] Lennarz also proposes that N_{535} may be involved directly in some function of Stt3p, which may be lost when N_{535} is glycosylated.[133] It should be noted that the *C. jejuni* pglB enzyme has a potential glycosylation site (N_{534}QS) that occurs after predicted catalytic site W_{457}WDYG.[133]

1.3.1.14 The Role of N-glycosylation

The single most important role of *N*-glycosylation is to promote proper folding of newly synthesized peptides in the ER.[62] It has been proposed that eukaryotic *N*-glycosylation evolved to a mechanism in which the protein substrate is presented to the

machinery in a flexible form before folding.[48] This form of glycosylation appears to be a eukaryotic adaptation of glycan function that allows cells to produce and secrete larger and more complex proteins at higher levels.[62] This role explains why *N*-linked glycosylation must occur co-translationally in the ER, before the folding process begins.[62] *N*-linked glycans have a direct effect on the folding process.[156, 157] Peptide studies indicate that although *N*-linked glycosylation does not induce permanent secondary structure, glycans alter the conformational preferences close to the gycosylation site, resulting in a more compact conformation.[157] For example, a truncated *N*-linked glycan GlcNAcβ1-4GlcNAc, the disaccharide chitobiose, tends to induce a compact β turn, and a total of one third of *N*-linked glycans occur in such locations.[62, 158, 159] A large polar saccharide unit is likely to affect the folding process locally by orienting the peptide segment towards the surface of protein domains.[62] When multiple sites on a given protein are modified, protein folding may be compromised even though none of the glycans are needed individually for the future function of the protein, thus, glycans appear to have a "global" effect as well as a local effect on protein folding.[62]

When *N*-glycosylation is inhibited, the most commonly observed effect is the generation of misfolded aggregated proteins that fail to reach a functional state, although dependence on *N*-glycosylation for folding varies on a case-by-case basis.[62] For example, during inhibition, some protein folding becomes temperature sensitive, while other proteins are only partially dependent on *N*-glycosylation, and some are dependent on *N*-glycosylation in one cell type but not others.[62] Also, some glycosylation sites are more important than others in a protein, which implies local effects on protein folding.[62] Refolding studies carried out *in vitro* using glycosylated and non-glycosylated versions of the same protein confirm that oligosaccharide moieties have a positive effect on the folding process.[156, 160]

Glycans are usually not essential for maintaining the folded structure, once folding is complete.[156, 161, 162] Although the presence of glycans increases protein stability, the effect is rather small. Most glycans can be removed after folding without major effects on protein structure/conformation.[62, 156, 157] Glycans keep the protein in solution and thus mimic roles played by molecular chaperones.[163] This former system would, of course, be more efficient for the cell. Instead of having a molecular chaperone for each protein, or even each class of difficult proteins, *N*-glycosylation can be considered as a generalized quality control system, which tags the protein as needing special "folding" attention. This system localizes potential problematic or difficult proteins to an intracellular compartment, where they can be scrutinized by cellular machinery. If they pass this molecular scrutiny, they will be sent through the secretory pathway, if they fail, these proteins will be refolded or destroyed to prevent them from causing problems in the cell.

1.3.1.15 The Calnexin-Calreticulin Cycle

Calnexin and calreticulin comprise a unique chaperone system found in the ER of nearly all eukaryotes.[131, 164, 165] Glycans are the "admission ticket" to this chaperone cycle.[62] Calnexin (CNX) is a membrane bound protein, while calreticulin (CRT) is a soluble chaperone.[131] These lectins, CNX and CRT, bind transiently to virtually all newly synthesized glycoproteins.[166, 167] Both chaperones interact with the glycans of the glycoprotein substrate after they have been trimmed by ER glucosidases I and II to the monoglucosylated form ($Glc_1Man_{9\text{-}6}GlcNAc_2$). However, if there is a glycosylation site within the first 50 amino acids, chaperone interaction may begin co-translationally.[62] After translation, lectin binding may continue for a period ranging from a few minutes to

several hours depending on the rate of folding, and these protein interactions slow down the rate of folding, but increase efficiency.[62]

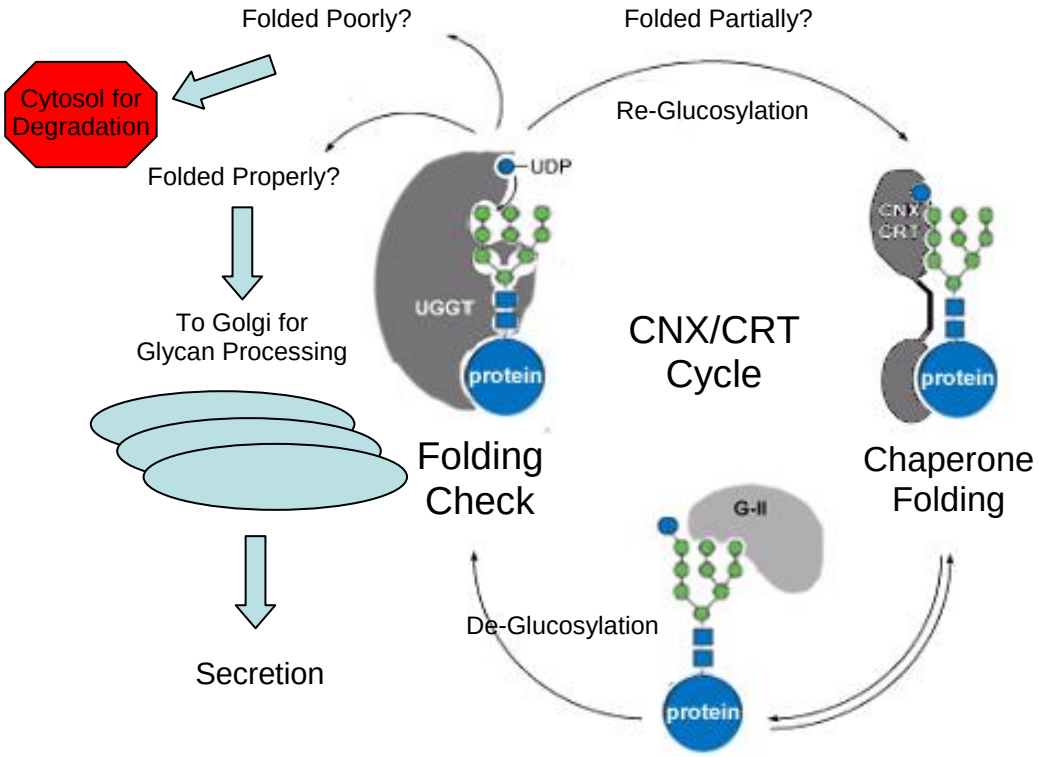
Both CNX and CRT are asymmetric molecules with a long, curved, hydrophilic peptide arm formed by the P-domain.[168-170] The P domain interacts with the ER cochaperone ERp57, or it may form a region that protects the bound substrate molecule during folding, but this does not form the lectin binding site.[62] CNX/CRT forms a complex with ERp57, a thiol oxidoreductase of protein disulfide isomerase (PDI), and short lived disulfide bonds form between CNX/CRT and substrate glycoproteins.[171] These disulfide bonds are transiently formed in oxidation and isomerization reactions, and lead directly to the formation of correctly paired disulfide bonds.[172]

1.3.1.16 UGGT Folding Sensor

Association of glycoproteins with CNX/CRT involves a binding and release cycle driven by the opposing actions of two soluble ER enzymes, UDP-glucose:glycoprotein glucosyltransferase (UGGT) and glucosidase II.[62] UGGT is a soluble 170 kDa glycoprotein that is ubiquitously expressed in the lumen of most eukaryotic cells.[173-176] UGGT is comprised of a large *N*-terminal folding sensor and a smaller *C*-terminal domain (catalytic).[177] UGGT adds a glucose residue to high mannose glycans and strongly recognizes the core pentasaccharide, $Man_3GlcNAc_2$, region of high mannose type glycoprotein oligosaccharide, suggesting its capacity to accept a wide range of *N*-glycan structures.[178] UGGT is a folding sensor, and only reglucosylates an incompletely folded glycoprotein.[165, 179, 180 181] Previous studies suggest that UGGT senses exposed hydrophobic patches of unfolded polypeptides proximal to the glycosylation site.[182-187] UGGT favors the presence of a hydrophobic substituent in proximity to the reducing end

of high-mannose glycans. As revealed through comparison of various acceptor substrates, the activity is of these acceptor substrates is governed by the hydrophobicity and orientation of the aglycone.[178] After glucosylation, the protein is redirected to folding chaperones, and the glucose residue is then removed. [62] Proteins that are folded properly are transferred to the Golgi where they undergo mannose trimming and elongation to one of three types of glycan structures as shown in figure 1.10.[97] If the protein is still not folded, when UGGT binds, the protein will be reglucosylated, and redirected to the folding chaperones.

1.9 CNX/CRT and UGGT Protein Folding Quality Control Cycle



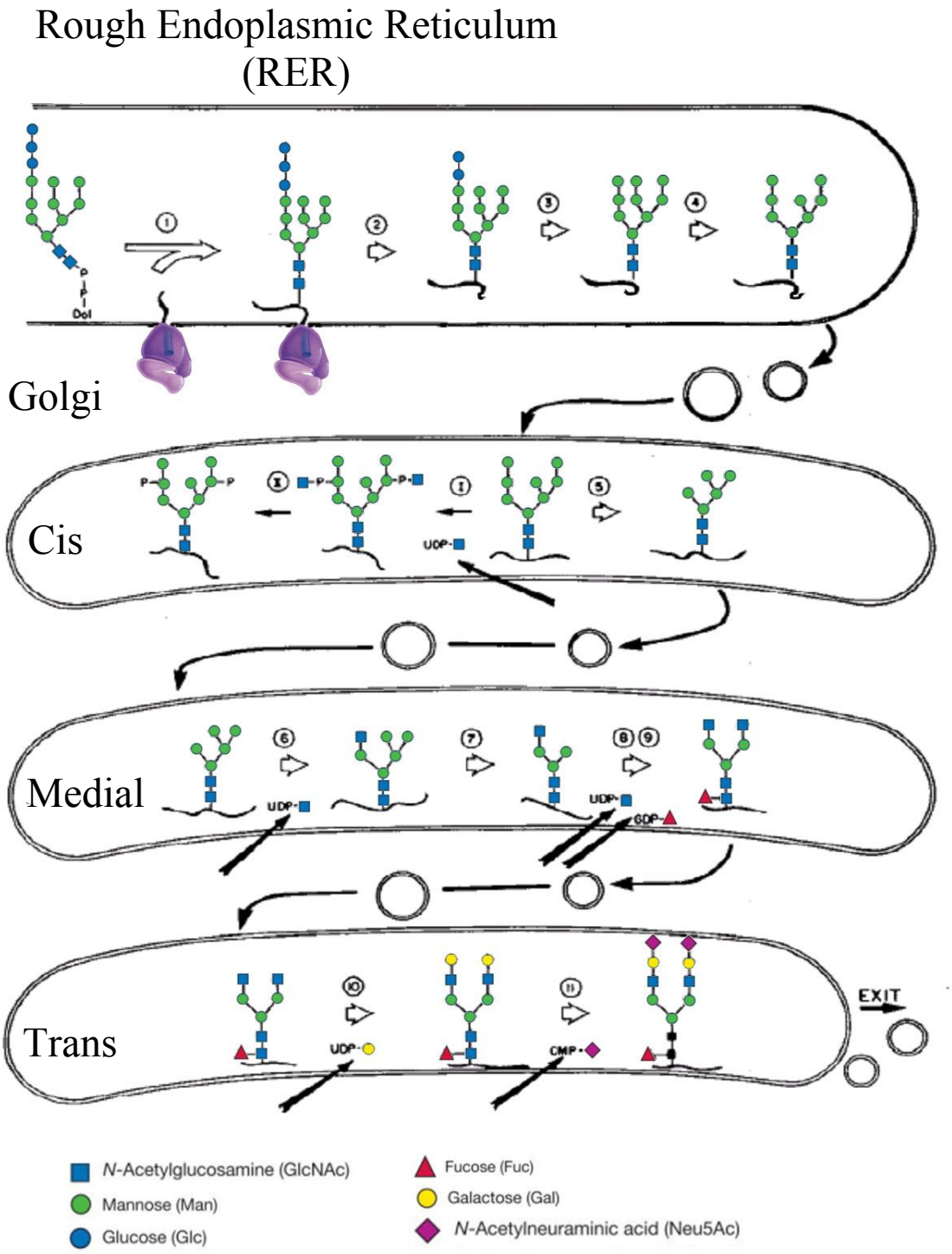
CNX/CRT appear to be essential *in vivo* because transgenic mice devoid of CRT die on embryonic day 18.[188] The essential nature of this cycle has also been shown through the lethal outcome of inherited glucosidase I deficiency involving a neonate born with severe generalized hypotonia and dysmorphic features.[189] These defects are likely the result of a shutdown of the CNX/CRT cycle.[62]

1.3.1.17 ERAD- The Fate of Misfolded Proteins

Mannose trimming at B and C arms of the glycan seems to function as a regulation system, in which accumulation of such glycans attenuates the glycoprotein entry into CNX/CRT cycle, consequently delivers GP to the ER associated degradation (ERAD) process.[190] When a glycoprotein is not folded properly, ER mannosidase I removes an α-1,2-linked mannose residue from the α-1,3 branch of the core oligosaccharide, which targets the resulting structure $Glc_{0-3}Man_8GlcNAc_2$ to ERAD.[62] If a glycoprotein is not folded properly it is targeted to ERAD, in which the misfolded protein identified via an unknown mechanism, exported to the cytosol, subsequently ubiquitinated, and finally degraded by the 26s proteasome.[191-193] The mannosidase responsible for the mannose trimming activity is quite slow, which may protect proteins against premature degradation.[161, 194, 195]

It is apparent that Man_8 structures serve as part of the signal needed for ERAD, although mannose trimming alone is not sufficient for ERAD, as correctly folded glycoproteins are trimmed of mannose residues before leaving the ER.[62] The ERAD model for degradation is not completely clear. In fact, multiple parallel pathways leading to degradation exist, and mannose trimming has been identified as one of the most conspicuous sorting criteria.[62]

1.10 The Fate of Properly Folded *N*-Glycans



1.3.2 MUCIN TYPE *O*-GLYCOSYLATION

Mucins are defined as cell surface or secreted glycoproteins with large numbers of clustered O-glycans.[196] Mucin-type *O*-glycosylation is initiated by the addition of a single GalNAc residue to the hydroxyl side chain of the fully folded peptide in the Golgi apparatus.[86] *O*-glycosylation is mainly a post-translational, post-folding event, thus only exposed Ser/Thr are expected to be glycosylated.[86] This type of glycosylation commonly occurs in Ser/Thr rich protein domains, but there is no known glycosylation consensus sequence, as observed in the case of *N*-glycosylation.[86] *O*-glycans are typically less branched than *N*-glycans, and commonly have biantennary structures.[196] It should be noted that not all *O*-glycans are mucin-type. Some proteins contain relatively few scattered *O*-glycans that have either short chains with a few residues, or elongated bi-antennary structures.[196] Although, in mammals, mucin type *O*-linked glycans on secreted membrane-bound proteins are the most common type of *O*-linked glycans that occur.[86] Mucin expression is enriched in epithelial cells specialized for mucus production and residing at the interface with the external environment. [196]

There are 8 basic or "core structures" for mucins, which begin with an α-linked attachment of an *N*-acetylgalactosamine (GalNAc) to the hydroxyl side chain of a Ser or Thr.[86] After the initial GalNAc attachment, glycans can be elongated to more complex structures, which can be terminated by the addition of sialic acid or fucosylation.[86, 196] This elongation process is highly regulated and highly complex. [86] The initial GalNAc additions vary due to differences in tissue specific expression of proteins. Site specificity has been found to be tissue specific by statistical analysis of large numbers of known attachment sites, as there are different GalNAc transferases with overlapping, but

different specificities, and these proteins have different tissue-specific expression patterns.[197, 198]

.

1.3.2.1 The Role of Mucin-type O-glycosylation on Protein Structures

Mucin-type *O*-glycosylation is commonly found on β-turns, and structures with extended conformation that often have high proline content.[86] By steric hindrance and repulsion between sialic acids, multiple *O*-linked glycans confer an elongated structure to the peptide backbone, essential for the function of mucins, and distort the α helicity of peptides.[86, 199] Mucins may fulfill role as "molecular spacers", distancing proteins from their neighbors.[86] Mucin type *O*-glycans are often found in regions of proteins with low hydrophobicity and no large side chains, such as tryptophan, or phenylalanine present. [86] *O*-glycans assist in the maintenance of secondary and tertiary structure of glycoproteins, and may play a role in the stabilization of quaternary structure of glycoprotein complexes.[86] Their roles in maintaining the three-dimensional structure of glycoproteins are accomplished through intermolecular interactions.[86] Finally, alternative protein splicing confers an additional regulation level for the properties and activity of different glycoproteins.[86]

1.3.2.2 The Signal Sequence for Mucin-type O-glycosylation

There is a primary sequence preference for mucin type *O*-glycosylation that is due to regional specificity.[86] The sequence preference is different for serine or threonine, as threonine appears to be glycosylated more efficiently than serine.[86] In the 8 positions preceding and following the *O*-glycosylated threonine, many serine, threonine, proline and valine residues occur, with proline or other uncharged residues commonly occurring

at positions -1 and +3.[86, 196] In the case of Serine *O*-glycosylation, a specificity window of 15 or 39 amino acids was identified and in many cases proline, threonine or serine was present at positions -9, -8 and +1, with respect to the glycan.[86] Cysteine has not been found at positions -2 to +2, and tryptophan is never at positions -1 and +1.[86] Methionine, aspartate and asparagine are rarely seen in proximal to mucin-type *O*-glycosylation sites, nor are positively charged residues seen in the immediate vicinity of the attachment site.[86] Likewise, a negatively charged residue at position -1 has never been documented.[86] Regardless of whether the glycosylation occurs on serine or threonine; alanine, serine and threonine are commonly found adjacent to the glycosylated residue. [86, 196] These rules for *O*-glycan attachment were established by comparison of different sites on known glycoproteins, without reference to the tissue-specificity. [86]

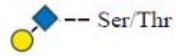
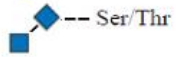
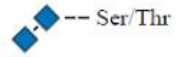
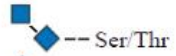
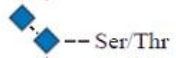
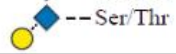
1.3.2.3 Mucin-type O-glycosylation Core Structures

The majority of *O*-glycans produced *in vivo* are comprised of core structures 1-4 with the Core 2 subtype most abundant. Most *O*-glycans contain the Core 1 subtype structure, which is formed by the addition of galactose (Gal) in a β1–3 linkage to the initial GalNAc moiety (Galβ1–3GalNAcα-Ser/Thr).[196] Core 2-type *O*-glycans are generated by addition of GlcNAc to the GalNAc in a β1–6 linkage forming (GlcNAcβ1–6GalNAcα-Ser/Thr) and contain the Core 1 subtype within the Core 2.[196] The Core 2 *O*-glycan may be elongated to either a mono- or biantennary form with the presence of multiple lactosamine (Galβ1-4GlcNAc) units and terminal linkages of fucose and sialic acid.[196] In myeloid cell types, Core 2 *O*-glycans appear to function as a scaffold for the production of selectin ligands that act in regulating inflammation.[196] The absence of the Core 2 GlcNAcT enzyme which produces Core 2 *O*-glycans results in a partial deficiency

in E-, L-, and P-selectin ligand biosynthesis.[196] Core 3 *O*-glycans can serve as building blocks for the formation of bi-antennary *O*-glycans.[196] Core 3 is found in few tissue types and may compete against Core 1 formation as both Core 1 and Core 3 require the same substrate.[196] Core 3 and Core 4 are mostly localized to the intestinal tract.[196]

.

1.11 Mucin-type *O*-Glycan Core Structures 1-8

Core	Structure	
Core 1	Galβ1-3GalNAcα1-Ser/Thr	
Core 2	Galβ1-3(GlcNAcβ1-6)GalNAcα1-Ser/Thr	
Core 3	GlcNAcβ1-3GalNAcα1-Ser/Thr	
Core 4	GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα1-Ser/Thr	
Core 5	GalNAcα1-3GalNAcα1-Ser/Thr	
Core 6	GlcNAcβ1-6GalNAcα1-Ser/Thr	
Core 7	GalNAcα1-6GalNAcα1-Ser/Thr	
Core 8	Galα1-3GalNAcα1-Ser/Thr	

Core 5 *O*-glycans (GalNAcα1–3,GalNAcα-Ser/Thr) have been found on glycoproteins from several vertebrate species. This structure exists on some human adenocarcinomas and on cells comprising the embryonic gut.[196] Core 6 *O*-glycan formation (GlcNAcβ1-6GalNAcα-Ser/Thr) has been reported on human embryonic gut and mucins from ovarian cysts, although it is possible that the Core 6 subtype is a Core 2 degradation product.[196] Core 7 O-glycans (GalNAcα1-6GalNAcα-Ser/Thr) were found in a study of *O*-glycans from bovine submaxillary sources.[196] In addition, linkage of fucose (α1–2) to the GalNAcα-Ser/Thr has been observed.[196] Further *O*-glycan diversification

continue to grow with the elucidation of more novel core subtypes, although most of these new subtypes are found at relatively low levels and are most likely segregated to specific tissue or cell types.[196] After core formation, these chains are elongated by the sequential addition of sugars from NDP-sugar donors and exported though the secretory pathway.

1.3.2.4 Role of Mucin-type O-glycosylation

Upon secretion, mucins generally remain at the surface of epithelial cells and are capable of providing a gelation function by retaining water and by their entanglement through intra- or intermolecular disulfide linkages. [196] Thus, mucin type *O*-glycosylation may be crosslinked by disulfide bonding in solution to promote gelation. Lack of mucins in the digestive tract may result in gastric ulcers.[196] Salivary mucins also provide lubrication for epithelial surfaces and may modulate the infectious nature of oral microbes, thus a lack of salivary gland function may yield abnormal mucosal surfaces with chronic low-level inflammation.[196] Salivary mucins also adhere to tooth enamel, thereby protecting against demineralization by organic acids, while others can agglutinate certain oral streptococcal strains.[196]

O-Linked glycosylation modulates the aggregation of glycoproteins, maintains protein stability, and confers protease and heat resistance.[86] Some mucin-type O-glycans act as temperature barriers against low temperatures to protect living tissues from freezing.[196] Such "anti-freeze" glycoproteins have been characterized and shown to inhibit the formation of "nucleation centers" at temperatures below the freezing point of water.[196] In these anti-freeze proteins, the peptide sequence "Ala-Ala-Thr" is repeated, in some cases more than 40 times, with each threonine residue carrying the *O*-glycan

structure Galβ1–3GalNAcα.[196] Chemical elimination of these *O*-glycans has been shown to abolish the antifreeze function of these proteins.[196]

The *O*-glycan role as "molecular spacer" was previously mentioned. They are important, especially for membrane bound receptors where the functional part of the receptor is oriented to the extracellular space because of the presence of glycans.[86] *O*-glycans have been reported to function in sperm binding to the egg. The mammalian egg coat (the zona pellucida) contains a large number of *O*-glycans, as well as some *N*-glycans. Removal of egg *N*-glycans by glycosidase treatment does not abolish sperm binding, but loss of *O*-glycans following mild alkali treatment ablates sperm binding.[196]

Involvement of alterations of *O*-linked glycosylation in health and disease is relevant for further medical research and disease treatment.[86] For example, *O*-glycosylation of collagen plays an important role in arthritis.[86] Blood coagulation factor IXa, a naturally occurring mutation (R94S) is associated with hemophilia B defect. The mutation introduces an O-glycosylation site, which in turn is responsible for a strong reduction in the activity of factor IXa.[86] Various pathogens can use these oligosaccharide structures found on cell surface *O*-glycans as receptors for binding and host infection.[196] Multimericity of the receptors and ligand multivalency can result in higher affinity binding, which is crucial for rotavirus binding to sialic acids.[86] In some cases, secreted mucins contain identical oligosaccharide structures at high densities and are capable of binding to pathogens, saturating pathogen receptors.[196] Thus, *O*-glycans on secreted mucins have been suggested to act as "decoys" in an organism's effort to evade pathogen infection, luring the pathogen in question to bind to the mucin rather than the host tissues.[196]

The vertebrate hematopoietic system utilizes *O*-glycans during both development and function.[196] *O*-Glycans play important roles in the recognition of other glycoproteins

such as selectin, ABO blood group antigens, MHC-complex.[86] Lymphocyte homing and the leukocyte inflammatory response are regulated by glycans that function as selectin ligands.[196] The role of *O*-glycans in immune response, selectin-mediated trafficking and cellular adhesion processes is quite extensive. For example, the involvement of selectin ligands in cancer cell metastasis has been proposed from studies revealing reduced metastasis in selectin-ligand-deficient systems.[196] Thus, it appears that specific *O*-glycan linkages have unique roles in various physiologic systems.[196]

O-Glycans are also essential for intermolecular interactions, and have important regulatory roles, including the modulation of the activity of enzymes and signaling molecules and in cellular glycoprotein expression and processing.[86] It is possible that *O*-glycan formation is under the control of glycosyltransferases that operate in protein-specific glycosylation reactions *in vivo*.[196] This implies that the function of some glycoproteins may be dependent on the *O*-glycan structure attached.

1.4 SELECTED TYPES OF MONOGLYCOSYLATION, *O*-GLCNACYLATION, *O*-MANNOSYLATION AND *C*-MANNOSYLATION

1.4.1 *O*-GlcNAcylation

1.4.1.1 O-GlcNAcylation Overview

It is now known that many proteins within cellular compartments are dynamically modified at their serine and threonine hydroxyl groups by the attachment of *O*-linked *N*-acetylglucosamine monosaccharides, which are dynamically attached in a β-linkage to proteins. *O*-GlcNAcylated proteins that have been identified are functionally diverse, however they share two common features: (1) they are also phosphorylated and (2) they form reversible multimeric complexes with other polypeptides or structures, and these

associations are often regulated by phosphorylation.[200] Thus far, the majority of the hundreds of *O*-GlcNAc-modified proteins detected by two-dimensional gel analyses of the nuclear and cytosolic fractions have not been identified. [201]

1.4.1.2 Discovery of O-GlcNAcylation

β-*O*-*N*-Acetylglucosamine (β-*O*-GlcNAc) addition to serine and threonine residues on proteins was first shown to be a major form of intracellular glycosylation in murine lymphocytes in 1984.[202] In 1986, the *O*-GlcNAc modification was found to be abundant in the cytoplasm and in virtually all subcellular organelles of rat liver, except for mitochondria.[203] These studies also found that the nucleoplasm contains the largest number of the *O*-GlcNAcylated proteins.[203] Rapidly, during the next two years, several different groups showed that nuclear pore proteins, which regulate trafficking of molecules into and out of the nucleus, are extensively *O*-GlcNAcylated.[201, 203] Since these early studies, *O*-GlcNAc has been reported to occur on a large number of proteins. [201] During the past several years, *O*-GlcNAc-modified proteins have been shown to be almost exclusively restricted to the cytoplasm and nucleus, except extracellular α-linked *O*-GlcNAc saccharides found in slime molds, and are abundant in virtually all eukaryotes examined, including protozoa and fungi, as well as in viruses.[204]

1.4.1.3 The Role of the O-GlcNAc Modification

It should be noted that β-*O*-GlcNAcylation is much more like phosphorylation than other forms of glycosylation, in that this form of protein post-translational modification is reversible and even reciprocal to phosphorylation, and dynamic.[200] Incredibly, β-O-GlcNAcylation has been shown to play a role in almost every phase of

cellular growth and normal cellular processes.[200] The removal of β-OGT from the cell results in cell death, embryonic lethality in murine models and reduced cell viability.[205] The cellular processes that β-OGT has been found to be involved in are too numerous to describe in detail, thus only a few general areas will be listed here.

The importance of the nuclear pore *O*-GlcNAcylated proteins in nuclear transport has been shown. Studies from several laboratories suggest a model in which the nuclear pore glycoproteins play a part in the initial peripheral binding of transport molecules.[201, 203] This is followed by the docking of the transported molecule at the center of the pore complex, and its translocation.[201, 203] Many different chromatin proteins, with a broad range of physical associations with chromatin, are *O*-GlcNAcylated.[203] RNA polymerase II and virtually every RNA polymerase II transcription factor examined to date contains *O*-GlcNAc. [201, 203, 206] Functions proposed for *O*-GlcNAcylation of transcription factors include nuclear transport, assembly into multimeric complexes, and regulation of phosphorylation.[201] Recent findings suggest that *O*-GlcNAc on transcription factors in fungi is important to transcription even in lower eukaryotes. [201, 203]

Many nuclear oncoproteins and tumor suppressors such as p53, c-Myc, v-Erb-a and the SV40-T antigen are *O*-GlcNAcylated.[201] Cytoskeletal bridging proteins, and other cytoskeletal proteins such as β-amyloid are *O*-GlcNAcylated, suggesting a role in Alzheimer's disease.[201] Cytokeratins, all three types of neurofilament proteins and microtubule associated proteins are modified by *O*-GlcNAc.[201] More recent research has demonstrated the role of this modification in glucose regulation and metabolic homeostasis at the cellular level.[201] Clearly, this modification is broad reaching and is involved in almost every aspect and function of cellular growth and development.

1.4.1.4 The O-GlcNAc Glycosyltransferase, β-OGT

The β-*O*-GlcNAc glycosyltransferase is highly conserved evolutionarily, with more than 85% similarity and >65% identity between the primary sequence of the enzyme derived from the nematode, *Caenorhabditis elegans*, and humans.[204, 207] Homologous open reading frames can be found in the public databases as far down the evolutionary scale as cyanobacteria.[204] The *N*-terminus of the protein has a region of multiple (11.5) tetratricopeptide repeats, or 23 anti-parallel α-helices, which create a superhelix, while the *C*-terminal domain contains the enzyme active site.[207] Tetratricopeptide repeats are 34 amino acid sequences that are known to mediate protein-protein interactions, which would enable the glycosyltransferase to achieve broad substrate specificity.[208] Instead of a single enzyme binding site, which would restrict the ability of β-OGT to bind to multiple partners, the TPR domain enables binding to many substrates without a dedicated binding site for each partner enzyme.[207] The TPR repeat region also modulates dimerization of β-OGT.[207]

It should also be noted that β-OGT, like the active component (Stt3p) of OT in *N*-glycosylation is apparently self-glycosylated, and in addition is phosphorylated as well.[209] Unlike Stt3p, an active site in the enzyme is not known, nor has a potential active site been proposed.

1.4.2 *O*-MANNOSYLATION

1.4.2.1 O-Mannosylation Overview

The *O*-mannose modification is a monosaccharide linkage in which mannose is transferred from Dol-P-man to a serine or threonine residue of a protein in an α linkage.[210] This modification occurs in the lumen of the ER and is catalyzed by two

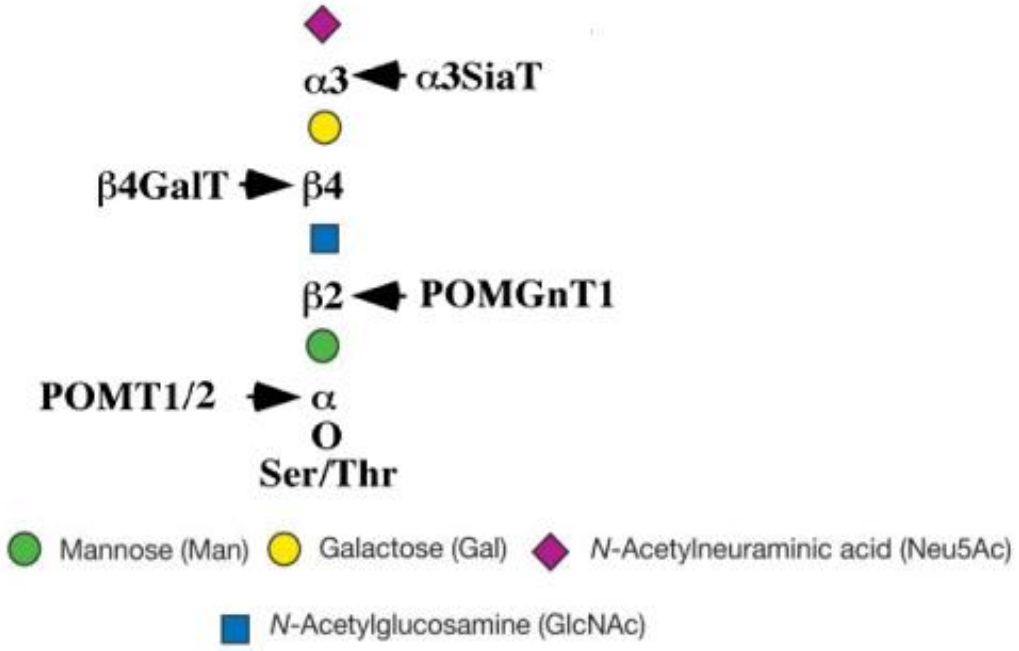
protein-*O*-mannosyltransferase proteins, POMT-1 and POMT-2.[95] This glycosylation event occurs co-translationally.[211] The initial glycosylation modification may be elongated by several additional mannose residues using a battery of mannosyltransferases that extend similar structures on *N*-linked chains in yeast.[210]

1.4.2.2 O-Mannosylation Discovery

O-αMan was identified in yeast in the 1950s by Margolis and coworkers, where *O*-mannosylated glycoproteins are abundant in the yeast cell wall.[212, 213] In fungi, all *O*-mannosyl glycans are neutral linear glycans consisting of 1-7 mannose units.[212] Other than fungi, *O*-linked mannose is also found in the skin collagen of a clam worm.[212]

O-αMan was identified in 1979 as the protein-linked sugar in rat brain proteoglycan.[214] In these studies, the glycan was released by mild base hydrolysis with simultaneous reduction of mannose to mannitol, thus the anomeric linkage could not be determined.[214] More recently, *O*-mannose-based glycans containing four to eight sugars have been isolated from total brain glycopeptides and in a tetrasaccharide derived from Ser/Thr-rich domains in bovine peripheral nerve α-dystroglycan.[215] This structurally important protein is found on Schwann cell membranes and forms a linkage with another membrane protein, β-dystroglycan, to laminin in the extracellular matrix.[210, 215] The oligosaccharide Sia2–3Galβ1–4GlcNAc-β1–2Man-Ser/Thr (figure 1.7) accounts for two thirds of the *O*-glycans of α-dystroglycan and is probably involved in the important α-dystroglycan-laminin binding, since sialidase prevents the interaction, but removal of *N*-glycans does not.[86] Similar glycans (Siaα2-3Galβ1-4GlcNAcβ1-3Man1-Ser/Thr) have been found in rat brain proteoglycans.[216, 217]

1.12 *O*-Mannose Based Glycan Structure



A small portion of total rat brain glycopeptide contains *O*-mannose-linked glycans and carries the HNK-1 antigenic epitope, an unusual 3-*O*-sulfated glucuronic acid at the nonreducing end of the glycan chain, which is implicated in neuronal cell adhesion.[217] In addition, dystroglycan isolated from other sources has been shown to contain a fucosylated, nonsialylated *O*-mannose-based structure with the Galβ1–4GlcNAcβ1–2Man linkage.[210] Quite recently, Smalheiser et al. have shown that dystroglycan from sheep brain carried the Neu5-Acα2-3Galβ1-4GlcNAcβ1-2Man, Galß1-4-GlcNAcβ1-2Man, and Galβ1-4(Fucα1-3)Glc-NAcβ1-2Man structures.[218] To date, α-dystroglycan has been found to be *O*-mannosylated in human, rabbit, sheep and bovine tissue.[212] Other neural proteins besides α-dystroglycan carrying these interesting *O*-mannose glycans have not been identified, but they are believed to be abundant since the estimated ratio of mannose-terminated to GalNAc-terminated *O*-glycans in total brain glycopeptides is

approximately 1:3.[210] Indeed, 30% of protein *O*-glycosylation in the brain are *O*-mannose linked glycans.[213]

1.4.2.3 The Cellular Role of O-mannosylation

Thus, in higher organisms, *O*-mannosylation may be more diverse and complex than expected and *O*-mannose based glycans play an important role in specific cell adhesion processes.[210] *O*-Mannosylation is an evolutionarily conserved protein modification and may be essential for muscle development in both vertebrates and invertebrates.[95] A homolog of POMT1, the 'rt' mutant in *Drosophila* causes reduced fertility and viability as well as defects of myogenesis.[95] It is postulated that besides its neurological function, *O*-mannosylation may be required for the integrity of muscle structures in mammalian cells.[212] Dystroglycan complex links the extracellular matrix to the cytoskeleton in a wide variety of tissues.[212] α-Dystroglycan is the cell surface component of the dystroglycan complex, and is heavily glycosylated. It is known to bind laminin and agrin in the basal lamina of muscle cells and Schwann cells.[212] Sialylated *O*-glycan is involved in the interaction with laminin, and these *O*-mannosylated glycans may mediate the binding to laminin in multiple eukaryotic tissues. Certainly, α-dystroglycan is an essential component of the dystrophin-glycoprotein complex which functions in linking the actin cytoskeleton to extracellular matrix in muscle and nervous tissue.[213]

In mammals, *O*-mannosylation may also play important roles in the cellular recognition step in development. Dystroglycan is required for normal embryonic development beyond the egg cylinder stage.[219] Thus, defective *O*-mannosylation may hinder growth beyond this developmental stage. Infection may also involve α-

dystroglycan. Two unrelated pathogenic agents, several members of the arenavirus family including the Lassa fever virus, and lymphocytic choriomeningitis virus and *Mycobacterium leprae*, the bacterium responsible for leprosy, bind to target cells through interactions with a common receptor, α-dystroglycan. These glycans are important for infection by disease-causing agents.[212] These data suggest that the α-dystroglycan glycans may possibly contribute to invasion by these causative pathogens, although details of the underlying mechanisms remain to be determined.[212]

Protein-*O*-mannosylation appears to be required for normal structure and function of α-dystroglycan in muscle and brain.[95] The absence of POMT-1 results in an autosomal recessive disorder characterized by congenital muscular dystrophy, lissencaphaly, and eye anomalies termed Walker-Warburg syndrome (WWS) (OMIM # 236670).[220, 221] In severe cases of WWS, children are affected from birth, with likely brain malformation and rarely live beyond infancy.[95] POMT1 mutations cause complete loss of *O*-mannosyl glycans in WWS.[95] Since these glycans play important roles in adhesion process, a defect in the biosynthesis of *O*-mannosyl glycans has a severe effect on cell migration and cell adhesion.[95] Patients with mutations in POMT1 have normal POMT2.[95] The fact that POMT1 and POMT2 mutations are mutually exclusive suggests that POMT2 may be essential for normal development. Thus POMT2 mutations may cause embryonic lethality.[95] POMT1 is highly expressed in fetal brain, testis and skeletal muscle, while POMT2 is highly expressed in testis.[96, 222] However POMT1 and 2 are normally expressed in all human tissues.[96, 222]

In non-mammalian eukaryotic species, other types of *O*-linked mannosylation occur, most notably in yeasts, where its function does not involve neural cell adhesion or multi-cellular developmental stages.[86] Instead, in yeasts and other unicellular eukaryotic organisms, *O*-mannosylation is essential for cell wall rigidity and integrity, and this

glycosylation event is vital for protein modification and protein transport.[223] Note that in higher organisms, protein transport is not dependent on *O*-mannosylation. *Candida albicans* deficiency in POMT leads to defects in multiple cellular functions including expression of virulence.[224]

1.4.2.4 Protein-O-Mannosyltransferase (ScPmt1/2, POMT1/2)

The enzymes responsible for *O*-mannosylation are two protein-*O*-mannosyltransferases, POMT-1 and POMT-2.[95] POMT-1 and POMT-2 are active only when co-expressed, which suggests that a heterodimeric complex is necessary for the activity of the enzymes.[95] This dimerization holds true for the yeast homolog ScPMT1p, which requires a heterodimeric complex with ScPMT2p for maximum transferase activity.[225] It has been postulated that dimerization of the two POMTs may form a physical pore, which is in turn involved in flipping the Dol-P-Man across the ER membrane.[46] Dimerization may facilitate substrate binding and define specificity, or even that the POMT1/2 complexes may provide higher efficiency transfer.[225]

Protein-*O*-mannosyltransferases are evolutionarily conserved from yeast to humans.[225] These POMTs are integral membrane proteins localized to the ER and predicted to have multiple transmembrane helices.[127, 225, 226] Yeast homologs are postulated to have no less than 7 transmembrane domains, including a large hydrophilic ER oriented segment flanked by five *N*-terminal and two *C*-terminal spanning domains.[225] The N-terminal faces the cytosol, while the *C*-terminus faces the lumen.[225] Deletion mutagenesis has identified the *N*-terminal third of the transferase as being essential for the formation of a functional ScPMT1p-ScPMT2p complex. Mutation of R138 has demonstrated that this residue is critical for dimerization of ScPmt1p/2p and

may preserve a local arrangement of the *N*-terminal region essential for mannosyltransferase activity.[225] Deletion of the central hydrophobic loop (Loop 5), where the Mannosyltransferase Inositol and Ryanodine receptor (MIR) domains are located, eliminates mannosyltransferase activity but not protein-protein interactions.[225] Secondary structure predictions of loop 5 predict alternating α and β secondary structure.[225] Alignment of all PMT family members revealed that the central loop with MIR domains contains three highly conserved peptide motifs which can be considered as signature motifs of the PMT family.[225] Furthermore, point mutations of highly conserved residues within these motifs severely diminish transferase activity.[225] Clearly the invariant residues are essential for ScPMT1p function.[225] These proteins are purported to be inverting glycosyltransferases, using general acid/base catalysis, involving residues Glu78, Arg64 and Arg138.[225]

1.4.3 C-MANNOSYLATION

1.4.3.1 C-Mannosylation Overview

The most recently discovered protein glycosylation event, *C*-mannosylation, is the addition of a mannose residue in an α linkage to the C2 of the indole ring of tryptophan.[227] This glycosylation event is known to be protein catalyzed, and occurs in the lumen of the ER in most eukaryotic cells, but not in bacteria, yeasts or plants.[227] Since the discovery of *C*-mannosylation, 10 proteins have been experimentally identified to be *C*-mannosylated, although 336 proteins in the human proteome have been predicted to undergo modification thusly.[227, 228]

1.4.3.2 Discovery of C-mannosylation

Hofsteenge and co-workers first discovered the *C*-mannosylation modification on RNase2 from human urine in 1994.[229] At this time, it was unclear if the glycosylation was a chemical degradation signal, or a modification generated before the excretion process. It was subsequently determined that RNase2 was *C*-mannosylated upon secretion from mammalian cells.[228] During the next phase of this research, Hofsteenge and coworkers examined a variety of cultured cells including bacteria, yeast, insect and mammalian cells and established that higher eukaryotes such as mammalian cells and some insect cells exhibit *C*-mannosylation, whereas bacteria and yeast do not.[228] They also identified the *C*-mannosylation consensus motif WXXW, where the first tryptophan becomes *C*-mannosylated.[228] Replacement of the second tryptophan by other amino acids abolished or substantially reduced the level of *C*-mannosylation.[228] They went on to determine that this modification is protein catalyzed, and occurs in the lumen of the ER using the substrate Dol-P-Man as the biosynthetic donor for *C*-mannosylation.[126] Hofsteenge and coworkers went on to identify *C*-mannosylation on IL-12b, multiple complement proteins, properdin, thrombospondin, and the erythropoietin receptor.[227] During this time, this group also identified another *C*-mannosylation signal sequence termed a thrombospondin repeat or "TSR" (WXXWXXWXXC) in which multiple tryptophan residues were found to be *C*-mannosylated. They were unable to identify the enzymes responsible for catalyzing the *C*-mannosylation glycosylation reaction. Recently, the first viral protein, sGP from *Zaire ebolavirus* was shown to be *C*-mannosylated, and it was predicted that other viral proteins have the potential to possess this unique modification as well.[230] A novel marine natural product was also found to be *C*-mannosylated.[231] Taken together, these data indicate that this glycosylation event may have broad applications in both natural products biosynthesis and infectious disease research.

1.4.3.3 The Role of C-mannosylation

Due to the novelty of this protein post-translational modification, and the absence of knowledge of the enzyme, the function of this modification is unknown. Several theories have been proposed as to the role of *C*-mannosylation. It has been suggested that the function of the mannose residue is structural in nature, and is responsible for the orientation of the tryptophan residue in the protein.[232] In RNase2 for example, the mannosylation appears to be in a position in which it can stabilize the loop structure upon which it is located, and orient the indole ring of the tryptophan within the protein.[232] A simple function may be to increase the polarity of the hydrophobic residues in the aqueous extracellular environment. An interesting proposal involves the complement pathway, in which multiple proteins are *C*-mannosylated.[233] The biological consequence of the complement pathway is to activate mannose binding lectin (MBL), whose function is to bind terminal mannose residues in glycoproteins and pathogen membrane glycans.[234] It was proposed that perhaps the *C*-mannosylated residues could have some interaction with MBL, yet when this was examined, neither of these known mannose binding lectins, MBL or Concanavalin A (ConA), were able to bind *C*-mannosylated tryptophan.[234, 235] Yet another proposal, and perhaps the most practical, is that the *C*-mannosylation is a secretory message, directing soluble proteins for export out of the cell.[228] This is the most likely function of the modification, as the *O*-mannosylation secretory signal used in simple eukaryotic organisms has been co-opted and is now used for α-dystroglycan in higher eukaryotes. This would also provide the benefit of protein stability as *in vitro* galactosylation and sialylation of therapeutic glycoproteins have been suggested to increase serum half-life, and increasing the amount of glycosylation on proteins that are

used as therapeutics has been shown to increase and prolong *in vivo* activity.[236, 237] However, the role of *C*-mannosylation cannot be fully explored until the *C*-mannosyltransferase is identified.

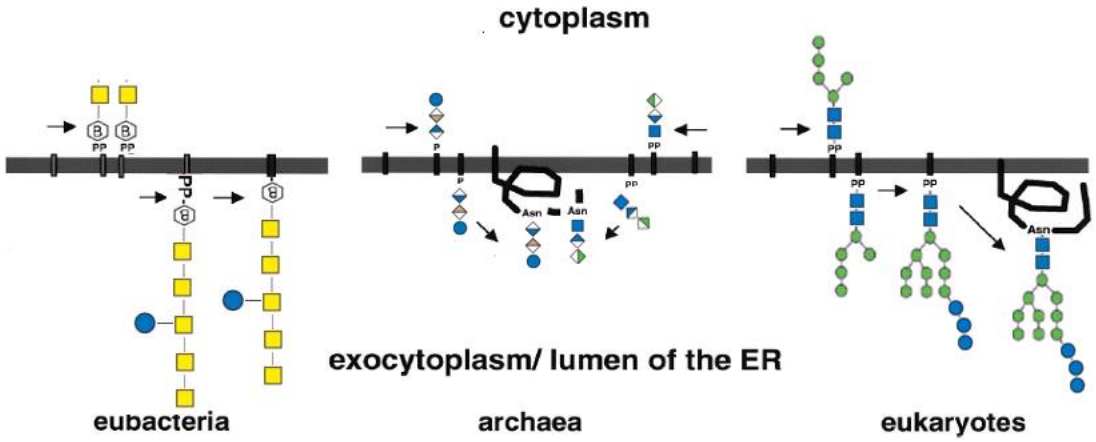
1.4.3.4 C-Mannosyltransferase

The focus of the second portion of this dissertation project is to find potential *C*-mannosyltransferase candidates, and to identify the *C*-mannosyltransferase. Taking into account what is known regarding the *C*-mannosylation, and other glycosylation events that occur in the ER, we can suggest several characteristics of the enzyme. First, the protein must have a Dol-P-man binding site, as is found on the N-terminus of POMT1 and POMT2.[225] This protein may also be a multi-pass transmembrane protein, like most of the OT complex involved in *N*-glycosylation as well as POMT1 and POMT2 in *O*-mannosylation.[133, 225] This protein may well be a complex of proteins, again as found in both *N*-glycosylation and *O*-mannosylation.[133, 225] Alternatively, this protein may be a homodimer, as observed with β-OGT.[209] It is also quite possible that, like β-OGT and Stt3p, the catalytic protein in the OT complex, the *C*-mannosyltransferase is *C*-mannosylated, thus the WXXW motif may be present in the protein. [137, 209] In addition to all these factors, the protein must also possess an ER targeting sequence at the *N*-terminus, which will be readily apparent upon protein sequence examination. This is a challenging task, and when completed will allow assay development to explore the protein function and potential cellular roles of the enzyme.

1.5 COMPARISON OF EUKARYOTIC AND PROKARYOTIC GLYCOSYLATION SYSTEMS

In summary, by comparing the similarities and differences between eukaryotic and prokaryotic glycosylation systems, an overall understanding of glycosylation can be reached. Extensive studies over years through protein identification, mechanistic elucidation and functional assignment have provided valuable knowledge of these pathways in multiple contexts. Comparison of disparate pathways also demonstrates striking similarities and conservation among glycosylation systems that may be extended to systems which have not been characterized. Here, each of the *O*-, *N*- and *C*- eukaryotic protein glycosylation systems is summarized and compared side by side with the prokaryotic systems.

1.13 *N*-Glycosylation Comparison of Prokaryotic and Eukaryotic Systems



As previously discussed, the most common modification of eukaryotic secretory proteins is *N*-glycosylation, which is well-characterized. The nucleotide-activated sugars are added sequentially to a dolichol pyrophosphate carrier and then translocated across the ER membrane. In prokaryotic *N*-glycosylation systems, the nucleotide-activated sugars are also added sequentially to an undecaprenyl pyrophosphate carrier and then

translocated across the cellular membrane. At this point in both the bacterial and eukaryotic systems, the glycan is a lipid-linked heptasaccharide containing one branching sugar. However, in eukaryotes, the heptasaccharide is processed to a tetradecasaccharide ($Glc_3Man_9GlcNAc_2$) prior to glycosyltransfer while prokaryotic glycans are not processed after this stage.

In the central step of the process, the OT complex transfers an assembled oligosaccharide from the lipid carrier to an asparagine residue in the emerging polypeptide. In eukaryotes, this process requires a multiprotein complex, while in prokaryotes only the Stt3 ortholog is needed. In eukaryotes, the covalently attached glycans are modified in the ER and Golgi complex by trimming, further glycosylation, sulfation and epimerization. In contrast, there is no evidence of further processing of the bacterial heptasaccharide. Interestingly, in archaea, the *N*-linked glycan synthesis proceeds through lipid-linked intermediates using both polyprenyl carriers and dolichol phosphate carriers on the extracellular cell surface. Further trimming reactions have also been described in archaea.[238]

The *N*-linked glycosylation sequence recognized by the OT, Asn-X-Ser/Thr, is conserved in eukaryotes, bacteria and archaea. It has been shown that the specificity of archaeal OT is not as strict as the eukaryotic and bacterial OTs, because replacement of the serine residue in one glycosylation site in *H. halobium* did not prevent *N*-glycosylation of the S-layer. It remains to be determined whether bacterial and archaeal glycosylation also show dependence on protein structural domains as is the case in eukaryotic systems, in which *N*-glycosylation is sensitive to the secondary structure of the protein. Also, in eukaryotes, the range of sugars that can be linked to Asn through a GlcNAc moiety is defined and limited as compared to those of bacteria and archaea.

These simple organisms are capable of synthesizing glycoproteins with multiple glycan structures and linkage types, resulting in a highly diverse assortment of glycoconjugates.

In eukaryotes, *O*-linked glycosylation occurs sequentially at Ser or Thr residues located in hydrophobic patches of the target protein. The product of one glycosyltransferase becomes the acceptor substrate for the next glycosyltransferase and nucleotide-activated sugar. There are no definitive sequons for *O*-glycosylation, although there are shared amino acid patterns among known glycosylation sites in proteins. This is similar to the *O*-linked process in bacteria. The process of *O*-linked glycosylation occurs mainly in the Golgi complex in eukaryotes, although *O*-mannosylation begins in the ER and *O*-GlcNAcylation occurs in the cytosol and nucleus. In archaea and bacteria, these glycosylation steps are much more reminiscent of *O*-GalNAc or *O*-Man based glycosylation, and occur at the plasma membrane or in the cytoplasm or at the interface between the cytoplasm and the surface appendages such as pili and flagella. It should be noted that six *O*-linked β-GlcNAc moieties were identified on the peritrichous flagella of *Listeria monocytogenes*.[239]

In bacteria, there is no known protein *C*-glycosylation. This is distinct from *N*- and *O*-glycosylation events which are common in prokaryotes. However, small molecule *C*-glycosylation in prokaryotes has been reported, and many of these compounds are aryl-*C*-glycosides.[73] The *C*-glycosylation of the aromatic ring in eukaryotic proteins suggests that there may be some similarity between the prokaryotic and eukaryotic *C*-glycosylation mechanism, although the enzymes may be divergent, as seen in bacterial and eukaryotic OTs involved in protein *N*-glycosylation.

It is worth noting that *O*- and *N*-glycosylation are conserved from the single-cell to the most complex organisms. However, *C*-mannosylation is found only in higher eukaryotes, and the roles of glycosylation change to a certain extent when progressing

from lower to higher order organisms. For example, *O*-glycans in archaea, bacteria and eukaryotes (*O*-GalNAc derived) are generally found on the cell surface and are responsible for cellular adhesion, ligand binding, cellular structure and cellular interactions. *O*-Mannosylation, which plays a role in muscle and neural function in higher eukaryotes has a similar function as proteins involved in flagellar function in lower organisms. The function of *N*-glycosylation in eukaryotes is to affect protein folding while in prokaryotes it seems to be related to protein complex formation. In both cases *N*-glycosylation may mediate proper protein functions. These comparisons are not perfect, as the function of *N*-glycosylation in archaea cannot be compared to those in bacteria and eukaryotes in the same fashion. Moreover, mono-glycosylation events such as *C*-mannosylation and *O*-GlcNAcylation have no equivalents in the single-celled world.

If a comparison could be made, one might correlate small molecule glycosylation of natural products to *C*-mannosylation of proteins. Small molecules can be glycosylated then exported from the cell. This is analogous to *C*-mannosylation of proteins to facilitate their export from the cell. Natural products are commonly used as targeted warheads to destroy bacteria invading the producing organisms. Interestingly, *C*-mannosylation is found on many proteins in the human immune system, and thus may be involved in protecting the body from foreign invaders. *C*-mannosylation is also found on viral proteins, suggesting that *C*-mannosylation may have evolved as a response to pathogens. Conversely, pathogens may have evolved this type of glycosylation to better evade the human immune system. The full array of *C*-mannosylated proteins is yet unknown, thus these suggestions are merely conjecture based on known data. Further investigation would certainly provide additional insight for better assessing the roles of *C*-mannosylation.

1.6 SUMMARY AND THESIS STATEMENT

This work involves two very different glycosylation projects. The first project, detailed in Chapters 2 and 3, deals with the elucidation of the glycosylation of the novel tetronolide natural product, kijanimicin. In Chapter 2, the biosynthesis of the deoxysugar TDP-L-digitoxose, from the kijanimicin pathway was investigated *in vitro*. The genes were identified from the cluster, cloned, expressed and the encoded products were purified. Activity was demonstrated for each enzyme in the pathway and in Chapter 3 the pathway was reconstructed *in vivo* using *Streptomyces lividans*. These strains of *S. lividans* were used to examine kijanimicin glycosyltransferase activity. We were able to demonstrate the activity of 3 of the 4 digitoxosyltransferases in the biosynthetic pathway, and propose a biosynthetic Scheme by which the tetrasaccharide chain is formed. We also identified two putative glycosidases with novel folds, and one glycosyltransferase that appears to have the unprecedented activity of attaching 2 if not 3 sugars in sequence.

In the second portion of this work, as described in Chapters 4 and 5, we attempted to identify the eukaryotic *C*-mannosyltransferase (CMT) enzyme and demonstrate its activity *in vitro* and *in vivo*. In Chapter 4, our efforts to locate the CMT are described. Through *in silico* analysis, putative *C*-mannosyltransferase genes were identified. These genes were expressed in *E. coli* and *S. cerevisiae*. However, gene expression was apparently toxic to *E. coli*. Although expression in *S. cerevisiae* was acceptable, extraction of the desired protein proved to be problematic. Thus, in Chapter 5, we describe our efforts to develop a CMT assay for use *in vitro* by expressing the putative CMT in insect cells. This approach was much more promising. We also attempted to knock down the putative CMT genes using shRNA. Our results showed that the putative CMT genes identified in this study were essential for cellular viability.

This work has contributed to the fields of both *C*-mannosylation and natural product glycosylation. We have elucidated the biosynthetic pathway of a novel deoxysugar, and identified potentially valuable tools for glycoengineering. These include a glycosyltransferase that appears to exhibit novel polymeric activity, and two glycosyltransferase proteins that are apparent glycosidases. Our attempts to identify CMT led to valuable information for the future development of specific assays for *C*-mannosylation. We also identified several promising protein candidates that are apparently essential for cellular viability in *H. sapiens*.

.

2 Prokaryotic Glycosylation: *Reconstruction of the Digitoxose Biosynthetic Pathway in Kijanimicin Biosynthesis*

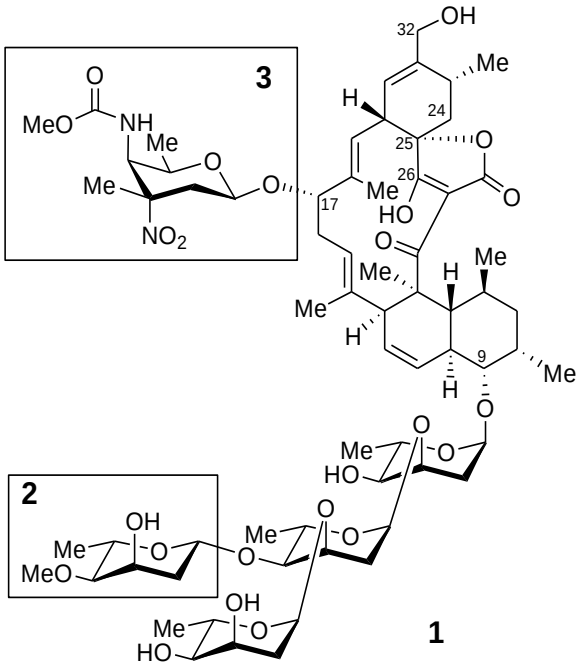
2.1 INTRODUCTION

Carbohydrates moieties decorate a wide variety of biologically active natural products that are used as therapeutic agents against diverse disease states including cancer, bacterial and fungal infections.[70] These natural products are secondary metabolites that result from complex multi-step biosynthetic pathways and are produced in a variety of organisms.[240] Although these compounds are diverse in structure and mode of action, their glycosylation status has been shown to be critical for the bioactivity of multiple classes of antibiotics including macrolides, peptides, and aminoglycosides, as well as natural-product anticancer, antiparasitic, and antifungal agents.[65, 70, 72, 240] Because removal of these sugar residues often results in the loss of biological activity of the parent compounds, altering native glycosylation patterns may vary the biological activity of their parent molecules [240, 241]

Addition of non-native carbohydrate biosynthetic pathways to an organism producing bioactive natural products, coupled with promiscuous glycosyltransferases has allowed the production of many novel products that are differentially glycosylated *in vivo*.[65, 240] To reconstruct the carbohydrate pathways which generate the sugar moieties *in vivo*, the genes involved must be identified and the relevant respective enzymes must be characterized. Also, application of glycosyltransferases to generate novel bioactive compounds with an altered carbohydrate pattern requires them to be promiscuous.[65] Mounting evidence from efforts devoted to studying this class of enzymes in the last few years indicates that many glycosyltransferases involved in the biosynthesis of secondary

metabolites, such as polyketide and non-ribosomal peptide antibiotics, have relaxed substrate specificity.[241] Polyketides such as macrolides, polyenes, tetracyclines and tetrononic acids, are compounds of great interest due to the diversity of structures and biological activities found in this subgroup of natural products.[240,242] Thus, increasing the available catalog of glycosyltransferases with known substrate flexibility would expand the biosynthetic possibilities of glycodiversification.

2.1 Kijanimicin A, a Secondary Metabolite Produced by *A. kijaniata*



The novel natural product kijanimicin, (see figure 2.1) was first isolated from *Actinomadura kijaniata* (ATCC 31588) in 1981.[243] It displays antibiotic activity against anaerobes such as *Propionibacterium acnes* and several strains of *Plasmodium* in murine

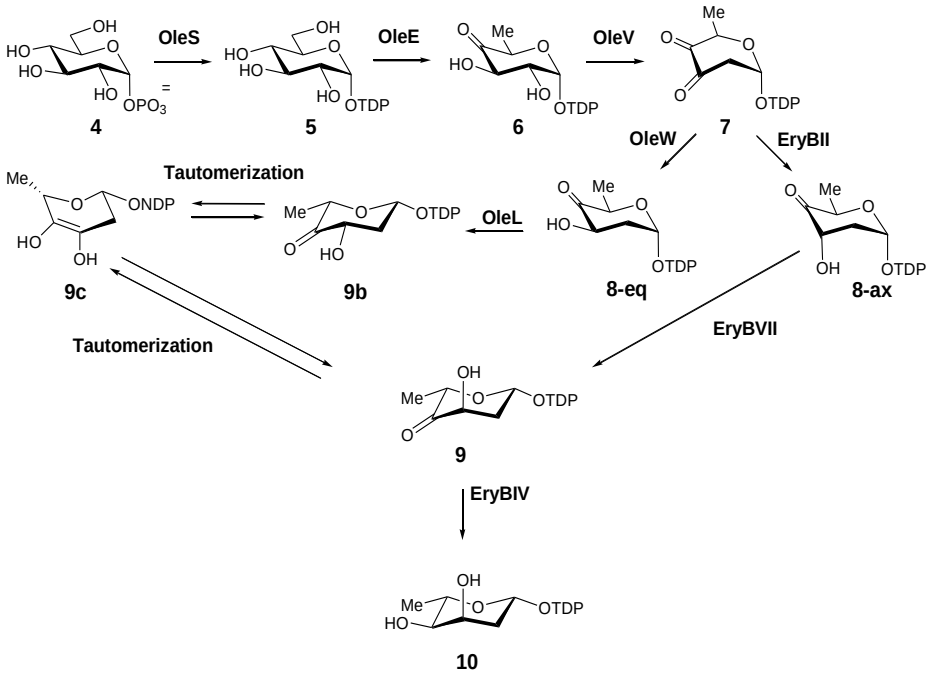
models.[243] The structure of the compound was subsequently determined by chemical degradation, X-ray crystallography and spectroscopic studies.[244-246] Kijanimicin is a polyketide, with a unique tetronic acid structure, containing a branched-chain tetrasaccharide moiety, consisting of three units of 2,6-dideoxy-α-L-hexopyranose (digitoxose) and one unit of 2,6-dideoxy-4-*O*-methyl-*β*-L-hexopyranose (4-*O*Me-digitoxose) (**2**), as well as the novel amino sugar kijanose (**3**).[246] Digitoxose and other 2,6-dideoxysugars, are ubiquitous in nature and are especially abundant in bacterial secondary metabolites, such as macrolides and aminoglycoside antibiotics.[241]

Besides kijanimicin, L-digitoxose is also found in many other bioactive compounds, such as in antibiotics jadomycin, several antifungal polyene macrolides from streptomycetes, and the tetrocarcin group of antitumor antibiotics from *Micromonospora chalcea.*[247-249] L-digitoxose has also recently been identified in new decalactone derivatives from the sea sponge *Latrunculia corticata.*[250] The D epimer of digitoxose is found as a structural component in cardiac glycosides (e.g., digitoxin), pregnane glycosides (e.g., cynaphyllosides isolated from the aerial parts of *Cynanchum aphyllum*) and apoptosis inducers (e.g., ammocidin, from *Saccharothix sp*. AJ9571).[251-253] D-digitoxose is also found in saccharomycins, a novel class of heptasaccharide antibiotics, which are active against multiple-drug resistant strains of bacteria.[254] The presence of digitoxose in natural products with varied biological functions makes this sugar an attractive building block to be used in combinatorial biosynthesis to glycosylate existing compounds.

It has been a longstanding goal in both synthetic chemistry and biology to build structural diversity by appending a wide variety of glycosides to a fixed aglycone scaffold. In 2002, Rohr and Mendez engineered a plasmid construct for their combinatorial biosynthesis work in which genes from the oleandomycin and

erythromycin gene clusters were used to generate thymidine diphosphate (TDP)-L-digitoxose *in vivo* via an unnatural pathway, involving non-enzymatic tautomerization between TDP- L-olivose and TDP- L-digitoxose.[255, 256]

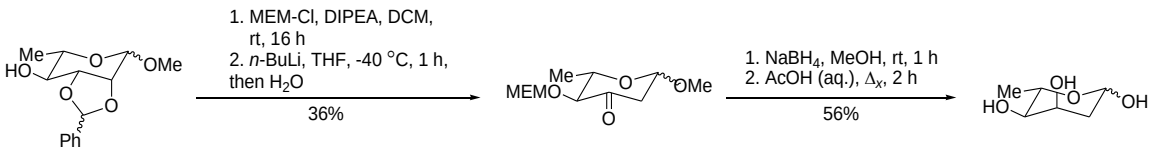
2.2 Combinatorial Biosynthesis of TDP-L-Digitoxose (from Mendez and Rohr)



In 2004, they engineered another plasmid construct that utilized the genes from the mithromycin, oleandomycin and erythromycin gene clusters to generate TDP-L-digitoxose *in vivo*. The key step of this preparation included the conversion of an intermediate with a 3-axial hydroxyl group, to the corresponding equatorial epimer by a 3,5-epimerase to generate TDP-L-digitoxose.[256] These biosynthetic pathways can be seen in figure 2.2. Although alternative biosynthetic routes exist for the synthesis of TDP-L-

digitoxose, elucidation of the native pathway would provide more efficient tools for the combinatorial endeavors. Therefore, the biosynthetic pathway from TDP-α-D-glucose to TDP-L-digitoxose was studied and fully elucidated in this work.

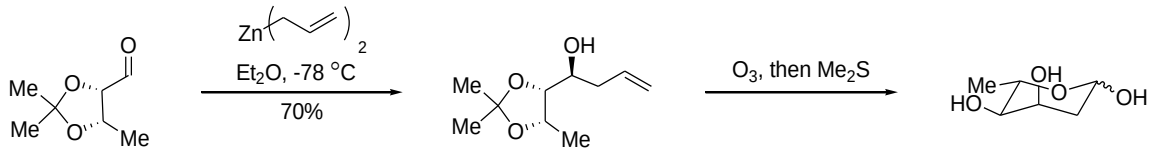
2.1 Synthesis of L-digitoxose by Brimacombe and coworkers [257]



Chemical syntheses of both D- and L-digitoxose are known.[257-263] Brimacombe and coworkers performed the synthesis of L-digitoxose using 2,3-*O*-benzylidene-α-L-rhamnopyranoside (Scheme 2.1). Synthesis commences by protecting the free hydroxyl with 2-methoxyethoxymethyl (MEM) group under standard conditions. Treatment of the protected sugar with *n*-butyllithium generates a methyl-2,6-dideoxy-α-L-erythro-hexopyranosid-3-ulose intermediate in 41% yield. This intermediate was further reduced with sodium borohydride to afford L-digitoxose and other isomers as minor products.[257]

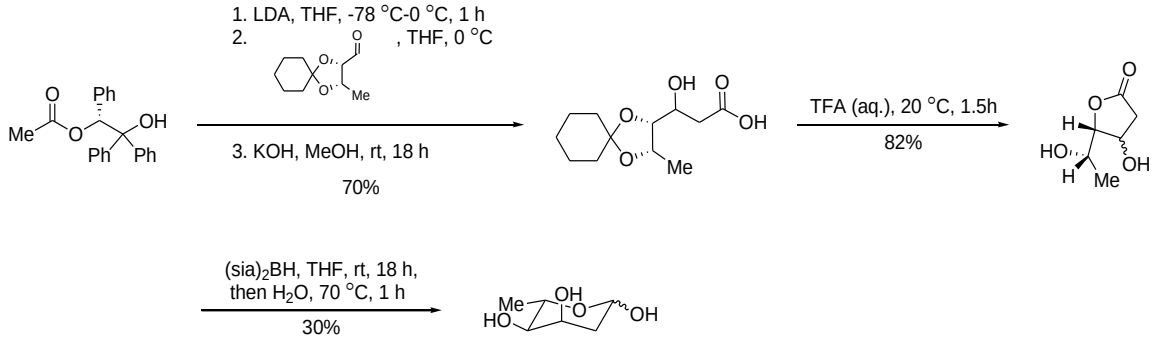
In 1982, Fronza et al. reported a three step synthesis of L-digitoxose in which 2,3-isopropylidene-4-deoxy-L-erythrose, the same chiral starting material used by Brimacombe and coworkers, was converted to an allylated carbinol by treatment with diallylzinc.[260] Subsequent exposure of the carbinol to ozonolysis followed by reductive workup affords the desired product L-digitoxose. Although the diallylzinc addition proceeds in moderate yield, the yield for the ozonolysis was not reported (scheme 2.2).

2.2 Synthesis of L-digitoxose by Fronza et al.



Braun and coworkers adopted an approach which could be used to synthesize both enantiomers of digitoxose, and synthesized the L form in approximately 4% yield (scheme 2.3).[261] This route uses chiral acetate synthesized from lactate and erythrose as the starting material to generate the required chirality. Both antipodes of the starting materials could be synthesized from commercially available materials. Aldol addition followed by ester hydrolysis gives a β-hydroxy acid, which cyclizes upon deprotection of the diol to give an intermediate lactone in good yield. This lactone, upon disamylborane reduction, gives digitoxose in an overall yield of 4% over 9 steps.

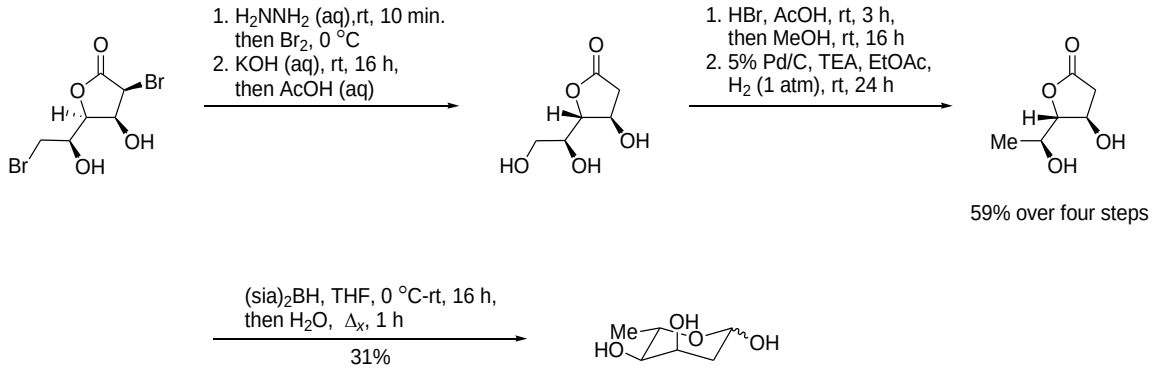
2.3 Synthesis of L-digitoxose by Braun and coworkers



Pederson generates the 2,6-dideoxy-L-*ribo*-hexose by first treating a L-*ribo*-lactone with hydrazine then bromine followed by basic hydrolysis to generate a 6-

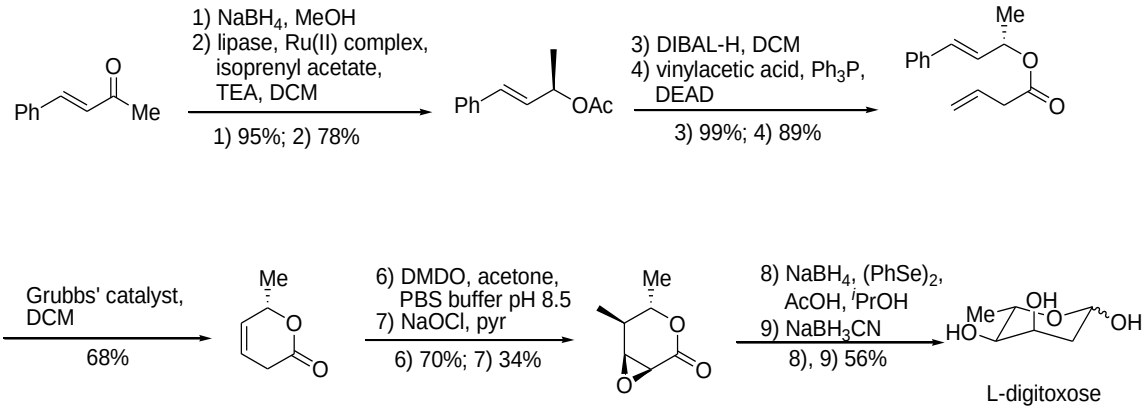
bromosugar with the desired stereochemistry.[263] As shown in scheme 2.4, the primary alcohol is removed via treatment of the diol with hydrobromic acid, followed by hydrogenolysis to furnish the 2,6-dideoxylactone (the same intermediate as Brimacombe's route), which could be isolated and purified as the diacetate. Subjecting the crude reaction product to disamylborane reduction generates the final product in 18% yield over 5 steps with only the final purification.[263]

2.4 Synthesis of L-digitoxose by Bock et al.



Wang and coworkers successfully synthesized L-digitoxose using ring closing metathesis on an allylic lactone, followed by further elaboration using redox chemistry to generate L-digitoxose.[262] An asymmetric enzyme-catalyzed isomerization prior to the ring-closing metathesis generates the desired enantiomer (see scheme 2.5). The yield is quite low; 6% overall yield over 9 steps.[262]

2.5 Synthesis of L-digitoxose by Wang and coworkers



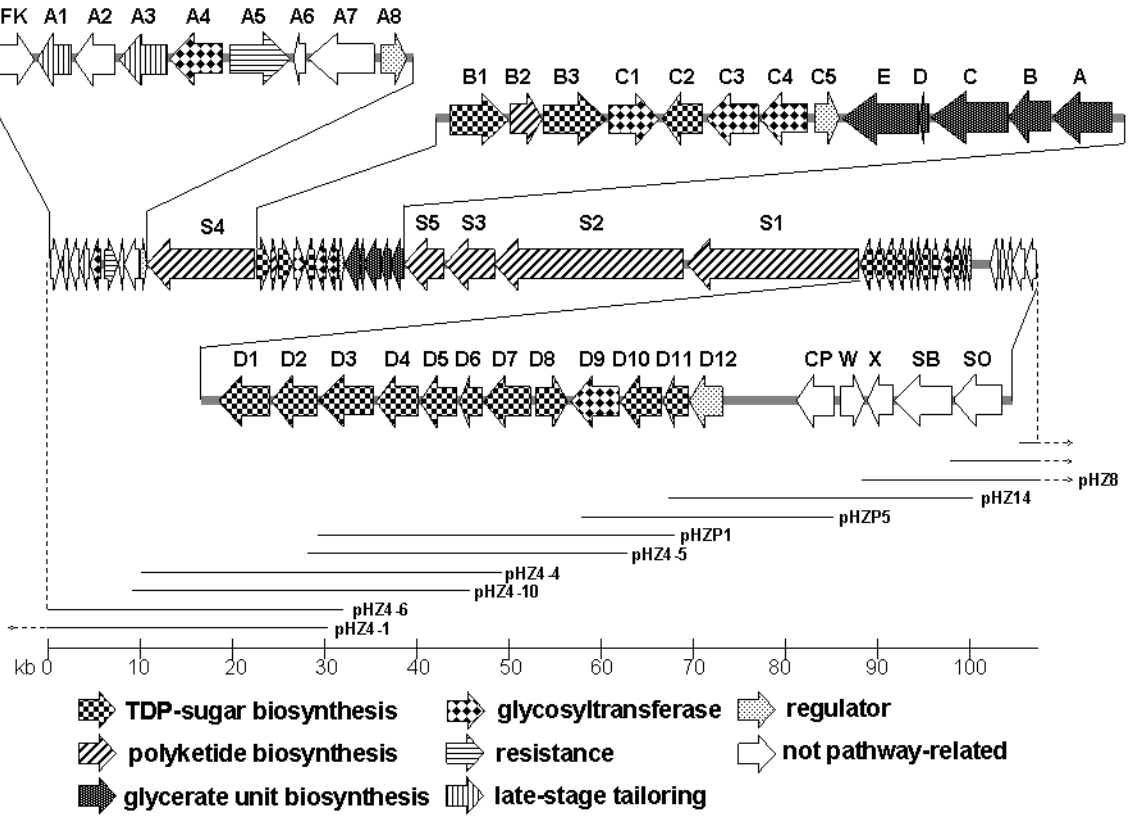
Overall yield: 6% over 9 steps

The advantages of elucidating the biosynthetic pathway of L-digitoxose are twofold. First, as illustrated above, the reported yields from chemical synthesis are relatively low. For biological applications, the sugar needs to be further activated through phosphorylation followed by TMP-morpholidate coupling at the C-1 position. The morpholidate coupling results in significant loss of compound, as the reaction affords both α and β anomers of the TDP-sugar.[264] Clearly, the biosynthetic approach has the advantage to be more stereoselective to generate the desired product than a chemical synthesis route. The second advantage of the biosynthetic approach is that it would also allow us to investigate the functions of the glycosyltransferases of the kijanimicin pathway *in vitro* and *in vivo*.

Thus, the first goal of this research project focuses on the determination of the enzymes involved in the biosynthesis of TDP-L-digitoxose. Once the pathway of digitoxose biosynthesis is established, we can investigate its subsequent attachment to the kijanolide aglycone during kijanimicin formation in *A. kijaniata*. This chapter will

describe the cloning and expression of the putative genes and the purification of the gene products proposed to be involved in the biosynthesis of TDP-L-digitoxose. To verify the functions of the enzymes involved in the biosynthesis of digitoxose, we have generated and characterized the carbohydrate intermediates produced at each step of the biosynthetic pathway.

2.3 The Kijanimicin Gene Cluster from *A. kijaniata*.



2.1.2 SEQUENCE ANALYSIS OF THE GENE CLUSTER

In order to understand the formation of kijanimicin, and more specifically the biosynthesis of digitoxose, we endeavored to reconstruct this deoxysugar biosynthetic pathway *in vitro*. We first subjected the gene cluster to *in silico* analysis to predict genes involved in the conversion of glucose-1-phosphate to TDP-L-digitoxose. Sequence alignments were undertaken by Dr. Hua Zhang, and the final gene cluster assembly is shown in figure 2.3. By comparing open reading frames in the gene cluster to known NDP-sugar biosynthetic enzymes candidate genes that may be involved in TDP-L-digitoxose biosynthesis were identified.

The amino acid sequences of KijD5 (298 aa) and KijD4 (375 aa) show homology respectively to glucose-1-phosphate thymidylyltransferases and TDP-D-glucose 4,6-dehydratases from various Streptomyces strains. The highest similarity of KijD5 is found with AAK83289.1 (59% identity and 73% similarity), the glucose-1-phosphate thymidylyltransferase in spinosyn biosynthesis from *Saccharopolyspora spinosa* and BAA84594.1 (59% identity and 71% similarity), the glucose-1-phosphate thymidylyltransferase localized in the avermectin biosynthetic cluster isolated from *Streptomyces avermitilis*.[265, 266] KijD4 showed high homologies towards AAK83290.1 (69% identity and 81% similarity), the TDP-D-glucose 4,6-dehydratase in spinosyn biosynthesis from *S. spinosa* and BAA84593 (68% identity and 82% similarity), the TDP-D-glucose 4,6-dehydratase localized in the avermectin biosynthetic cluster isolated from *S. avermitilis*.[265, 266] KijB1 (479 aa) shows the highest similarity to MtmV, TDP-D-4-keto-6-deoxyglucose-2,3-dehydratase localized in the mithramycin biosynthetic cluster from *Streptomyces argillaceus* (52% identity and 71% similarity).[267]

KijB1 and KijD10 (336 aa) possess 49.9% and 55% identities to the products of *gra-orf27* (CAA09648) and *gra-orf26* (CAA09647), in the granaticin biosynthetic

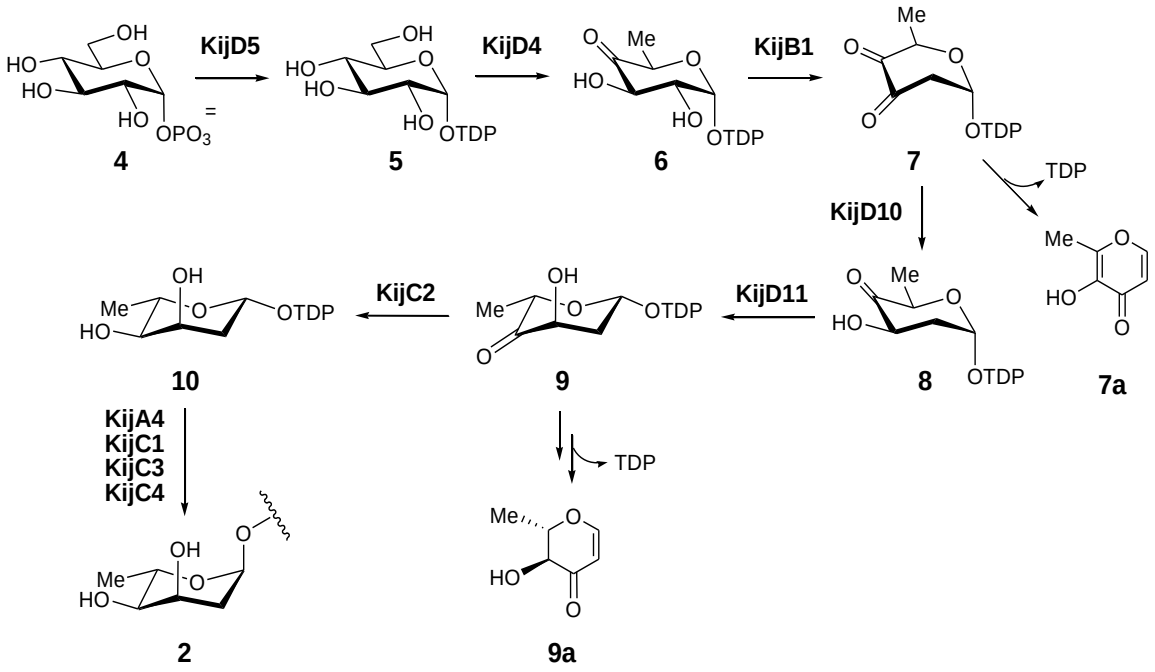
cluster, which have been proposed to encode a TDP-D-4-keto-6-deoxyglucose-2,3-dehydratase and TDP-D-3,4-diketo-2,6-dideoxyglucose-3-ketoreductase, respectively. The gene products of the granaticin cluster were biochemically characterized by Draeger et al.[268] Based on their analysis, it was predicted that KijD10 would be a TDP-D-3,4-diketo-2,6-dideoxyglucose-3-ketoreductase whose product has an equatorial hydroxyl group at the C-3 position (Figure 2.4, Compound **8**).

KijD11 (194 aa) exhibits significant homologies to 3,5- and 5-epimerase enzymes that act on a TDP-D-4-keto-6-deoxyhexose. The highest homologies occur with TylCII (AAD41825) (57% identity and 68% similarity) and the *gra-orf25* product (CAA09646, granaticin biosynthetic cluster) (52% identity and 68% similarity).[269] Significantly, the similarity of KijB1 to TylCII is a strong indication that this is a protein involved in digitoxose biosynthesis. The protein sequence of lipDig3, found in the α-lipomycin biosynthetic cluster, another tetronate class natural product, whose aglycon is decorated with digitoxose, shows 58% homology to TylCII.[270] The biosynthetic pathway of digitoxose should thus involve C-3 ketoreduction and C-5 epimerization yielding TDP-4-keto-2,6-dideoxyhexose, which is consistent with the presence of gra-orf26 and gra-orf27 in the kijanimicin biosynthetic cluster.[268, 271] Accordingly, KijD10 and KijD11 were proposed to mediate conversion of the KijB1 product (TDP-D-3,4-diketo-2,6 dideoxyglucose) to TDP-L-4-keto-2,6-dideoxyhexose.

The product of *KijC2* (391 aa) displays the highest homology to JadV, the (nucleotide diphosphate) NDP-4-keto-6-deoxyhexose-4-reductase found in the jadomycin biosynthetic cluster from *Streptomyces venezuelae* (46% identity and 58% similarity).[247] Jadomycin possesses an L-digitoxose unit attached to its anguicycline aglycone and gene disruption studies demonstrated that *jadV* is essential to the biogenesis of the L-

digitoxose moiety of jadomycin B.[247] The high similarity between KijC2 and JadV further suggests its involvement in the L-digitoxose pathway of kijanimicin biosynthesis.

2.4 TDP-L-Digitoxose Biosynthetic Pathway, adapted from Wang et al.



These sequence alignments led us to propose the TDP-L-digitoxose biosynthetic pathway as shown in Figure 2.4. Thus, in this chapter we cloned and heterologously expressed the genes *kijD5*, *kijD4*, *kijB1*, *kijD10*, *kijD11* and *kijC2*, for use in the *in vitro* assays to demonstrate their involvement in the biosynthetic pathway of TDP-L-digitoxose.

2.2 EXPERIMENTAL DETAILS

2.2.1 MATERIALS

2.2.1.1 Bacterial Strains

Escherichia coli (*E. coli*) strain DH5α and *Actinomadura kijaniata* SCC1256 (ATCC 31588) were purchased from Bethesda Research Laboratories (Gaithersburg, MD) and the American Type Culture Collection (Manassas, VA), respectively. The overexpression hosts *E. coli* BL21, BL21(DE3), BL21(DE3)PLysS and BL21 RosettaII(DE3) were purchased from Novagen (Madison, WI). *E. coli* XL-1 Blue MRF' used for cosmid library preparation was obtained from Stratagene (La Jolla, CA).

2.2.1.2 Biochemicals

Enzymes and the molecular weight standard used for the molecular cloning experiments were products of Invitrogen (Carlsbad, CA) or New England Biolabs (Beverly, MA). Restriction digestion enzymes, calf intestinal alkaline phosphatase (CIP), 100X bovine serum albumin (BSA), T4 ligase, and their respective buffers were products of New England Biolabs. Ni-NTA agarose resin was obtained from Qiagen (Valencia, CA). Growth media components were acquired from Becton Dickinson (Sparks, MD). Rabbit muscle pyruvate kinase was purchased from Sigma as a 400-800 unit/mg ammonium sulfate precipitate. This ammonium sulfate precipitate was dissolved in water to a concentration of 2500 units/mL, dialyzed against 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0, to remove ammonium sulfate, and stored at -80 °C. Antibiotics and chemicals such as isopropyl-β-D-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were

products of Sigma-Aldrich Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). Agarose for DNA electrophoresis was obtained from Fisher Scientific (Pittsburgh, PA). Bio-Gel P2 resin and all reagents for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were purchased from Bio-Rad (Hercules, CA), with the exception of the prestained protein molecular weight marker, which was ordered from New England Biolabs. Antibodies used for Western blotting procedures, monoclonal anti-polyhistidine clone His-I H1029 and anti-mouse IgG (alkaline phosphatase conjugate) A2429 were purchased from Sigma. For Western blot detection, nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) were both obtained from Promega.

2.2.1.3 Plasmids, Vectors and DNA Manipulation

Vector pET28b(+) was purchased from Novagen (Madison, WI). Cloning vector pANT841 was obtained from Dr. C. E. Hutchinson. Gigapack III Gold packaging extract, Plasmids SuperCos 1 and pBluescript II SK(+), used for cosmid library construction and cosmid fragment subcloning, respectively, were purchased from Stratagene (La Jolla, CA). DNA gel extraction and spin miniprep kits were obtained from Qiagen (Valencia, CA). DNA polymerase *pfu* was purchased from Stratagene. Oligonucleotide primers for cloning were prepared by Integrated DNA Technologies (Coralville, IA). The digoxigenin (DIG)-labeling and probing kit for colony hybridization assays was purchased from Boehringer Mannheim (Mannheim, Germany).

2.2.1.4 Instrumentation

PCR reactions were carried out using an Eppendorf Mastercycler gradient thermal cycler from Brinkman-Eppendorf (Westbury, NY). For agarose gel electrophoresis, a mini-sub-cell GT from BioRad (Richmond, CA) was used, powered by a Fisher FB-300 electrophoresis power source. After electrophoresis, DNA bands were visualized using a Fisher FBTIV-08 UV transilluminator. Cell disruption was performed with a Fisher 550 Sonic Dismembrator, equipped with a standard sonicator horn. For SDS-page gel electrophoresis, a mini PROTEAN II vertical system from BioRad (Richmond, CA) was used, equipped with an FB-300 or EC-1000-90 electrophoresis power source from Fisher, or E-C apparatus corporation, respectively. Mini-trans blot assembly apparatus used for Western blot, GelAir drying system for acrylamide gel preservation, and the necessary accessories and reagents were all products of BioRad. The nitrocellulose Hybond-C or Optitran BA-5 85 reinforced nitrocellulose membrane used for Western blot was from Amersham Bioscences or Whatman (Dussel, Germany), respectively.

To obtain pH values, a Corning 240 pH meter equipped with an Accumet electrode (Fisher) was used. Large scale centrifugation procedures were performed using an Avanti J-25 unit or a J-E instrument from Beckman (Arlington Heights, IL). Microscale centrifugation procedures were carried out using an Eppendorf 5415C from Brinkmann Instruments, (Westbury, NY). Ultraviolet-visible spectra were recorded using Beckman DU-650 spectrophotometer. Amicon YM-10 filtration products were purchased from Millipore (Billerica, MA). HPLC Separations used a Beckman 366 from Beckman Instruments (Fullerton, CA) equipped with a CarboPac PA1 HPLC column from Dionex (Sunnyvale, CA). ÄKTA FPLC Instrument was from Amersham Pharmacia Biosciences (now GE Healthcare). Mono-Q H/R 10/10 and 16/10 FPLC columns were from

Pharmacia (Uppsala, Sweden). DNA and biological compounds were dried using a Savant Speedvac SC100, equipped with a Savant refrigerated condensation trap RT100 and Gel Pump GP100. DNA concentrations were measured using a NanoDrop ND-1000 UV-Vis instrument from Thermo Fisher Scientific (formerly NanoDrop Technologies LLC).

NMR spectra were acquired on either a Varian Unity 300 or 500 MHz spectrometer. Sequencing of subclones was performed using a capillary-based AB 3700 DNA analyzer by the Core Facilities in the Institute of Cellular and Molecular Biology of the University of Texas at Austin. Mass spectra (low-resolution ESI) were obtained on a Finnegan LCQ ion trap mass spectrometer. Spectra were acquired by the Mass Spectrometry Core Facility in the Department of Chemistry and Biochemistry of the University at Texas at Austin or the College of Pharmacy of the University of Texas at Austin.

2.2.2 GENERAL PROCEDURES

All protein purification operations were carried out at 4 ºC. Protein concentrations were determined according to Bradford using bovine serum albumin (BSA) as the standard.[272] The relative molecular mass and purity of enzyme samples were determined using SDS-polyacrylamide gel electrophoresis as described by Laemmli.[273] NMR chemical shifts (δ in ppm) are given relative to that of deuterated water (D_2O, δ 4.65 for ^1H NMR) in the case of all NMR assays, with coupling constants reported in Hertz (Hz). The general methods and protocols for recombinant DNA manipulations were as described by Sambrook et al.[274]

2.2.3 EXPERIMENTAL PROCEDURES

2.2.3.1 Bacterial Strains and Growth Conditions

Actinomadura kijaniata SCC1256 (ATCC 31588) was maintained at 30 ºC on ISP medium 2. For genomic DNA extraction, a seed culture of *Actinomadura kijaniata* was grown in liquid V15 medium (seed media) (1% glucose, 0.2% peptone, 0.2% N-Z amine A, 1% beef extract, 0.2% $NaNO_3$, 0.1% K_2HPO_4, 0.054% $MgSO_4 \cdot 7H_2O$, 0.05% $CaCl_2 \cdot 2H_2O$, pH adjusted to 7.2 before autoclaving) for 24-30 h at 30 ºC, diluted into V15-P medium (20% w/v sucrose, 3% glucose, 1.5% bactosoytone, 0.3% glycine, 0.04% $CaCl_2 \cdot 2H_2O$, 0.1% $MgCl_2 \cdot 6H_2O$, pH adjusted to 7.2.), and grown for 40-48 h at 30 ºC.[275]

E. coli DH5α was used for routine cloning experiments. *E. coli* XL-1 Blue MRF' was used for cosmid manipulations. *E. coli* DH5α used for routine cloning experiments and BL21, BL21(DE3), BL21(DE3)pLysS, and BL21 RosettaII(DE3) for protein purification experiments were maintained on Luria-Bertani (LB) broth or LB-agar plates (1% NaCl, 1% tryptone, 0.5% yeast extract, pH 7.5, and in the case of plates, 2% agar). Culture media was supplemented with 100 µg/mL ampicillin or 50 µg/mL kanamycin with the addition of 35 µg/mL chloramphenicol in the case of BL21 RosettaII(DE3) and BL21(DE3)pLysS. To liquid cultures, volume of sterile 80% glycerol equal to 20% of the original culture volume was added and gently mixed to homogeneity. The mixture was aliquotted into sterile Eppendorf tubes, flash frozen in liquid N_2 and stored at -80 ºC until use. Glycerol stocks were used in a 1:20 dilution to inoculate fresh media for plasmid preparations.

2.2.3.2 Preparation of Competent Cells

The rubidium chloride (RbCl) method was used to make *E. coli* cells chemically competent.[274] A fresh colony of *E. coli* of the appropriate strain was used to inoculate 2 mL of Luria-Bertani (LB) (1% NaCl, 1% yeast extract, 0.5% tryptone, pH 7.5) broth. This culture was grown at 37 ºC with 250 RPM shaking. After overnight incubation, 1 mL of the culture was used to inoculate 100 mL of fresh LB media in a 1:100 dilution. The freshly inoculated culture was incubated at 37 ºC with 250 RPM shaking until the OD_{600} of the culture reached 0.4 to 0.5 absorbance units. At this time, the culture was transferred to a two chilled sterile 50 mL conical tubes and incubated on ice for 15 min. The cultures were centrifuged at 3,000 *g* for 10 min at 4 ºC. The supernatant was discarded and the cell pellets were gently resuspended in 33 mL of cold RF1 solution (100 mM RbCl, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$ 15% glycerol, pH 5.8, filter sterilized using 0.22 μm membrane). The cell suspension was incubated on ice for an additional 15 min and centrifuged again at 3,000 *g* for 10 min at 4 ºC. This RF1 washing step was repeated and the cells resuspended in 12.5 mL of RF2 solution (10 mM RbCl, 10 mM MOPS, 75 mM $CaCl_2$ 15% glycerol, pH 6.8, filter sterilized using 0.22 μm membrane). The cell suspension was incubated on ice for 10 min, and 50-100 μL of the mixture was dispensed into sterile eppendorf tubes. The aliquots were flash frozen in liquid N_2 and stored at -80 ºC until use.

2.2.3.3 Creation of a Cosmid Library from A. kijaniata

Cosmid DNA preparations were carried out by Dr. Hua Zhang. Standard genetic manipulations were performed as described by Sambrook et al.[274] Briefly, genomic DNA of *Actinomadura kijaniata* was isolated using a procedure derived from the SuperCos 1 Cosmid Vector Kit, with the following additional steps at the beginning of the procedure.

The cell pellet was washed with 10 mM EDTA followed by 3 h digestion at 30 ºC in Solution A (10 mM NaCl, 20 mM Tris•HCl, pH 8.0, 1 mM EDTA) containing lysozyme at 2.5 mg/mL final concentration. For the construction of *Actinomadura kijaniata* SCC1256 genomic library, chromosomal DNA was partially digested with *Sau*3AI and fragments of 20 to 40 kb were ligated into SuperCos 1 that was prepared by linearizing with *Xba*I, CIAP treatment, and subsequent digestion with *Bam*HI. Gigapack III Gold packaging extract was used to perform *in vitro* packaging, which was done according to the manufacturer's protocol. A 500 bp *kijd1* fragment labeled with digoxigenin (DIG) was used to probe the library as the starting point for chromosome walking. Cosmids pHZ-8, pHZ-14, pHZ-P1, pHZ-P5 and pHZ4-6 were digested with *Apa*I and *Bam*HI, and 1-5 kb fragments were subcloned into pBluescript II SK(+). DNA manipulations of *Actinomadura kijaniata* SCC1256 were performed as described by Kieser et al.[276] DNA sequencing data were assembled and analyzed using Vector NTI Suite program (version 6.0; InforMax, Inc.; Bethesda, MD). Protein sequences were analyzed with the NCBI BLAST server.

2.2.3.4 PCR Reactions

The general protocol for amplification of genes from Gram-positive bacteria is as follows. Primers were prepared by dilution with TE buffer (50 mM Tris, 1 mM EDTA, pH 8.0) to a concentration of 100 µM. This stock solution was further diluted to 20 µM with sterile water. Template DNA from the *A. kijaniata* genome was obtained from cosmids constructed by Dr. Hua Zhang. Genes *kijB1* and *kijC2* were amplified from cosmid pHZ4-6, and *kijD10* and *kijD11* were amplified from cosmid pHZ14. Polymerase chain reactions (PCR) consisted of the following: 72 µL of sterile water, 10 µL of 10X

pfu polymerase buffer, 10 µL of dimethylsulfoxide (DMSO), 10 µL of a 20 mM deoxyribonucleotidyl triphosphosphate (dNTP) solution, 2.5 µL of each primer (5′ and 3′) at 20 µM, 1 µL of template DNA, and 2 µL of *pfu* polymerase (5 units), for a 100 µL total reaction volume in an 0.5 mL thin walled PCR tube.

The PCR reactions were carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial denaturing step of 2 min at 95 ºC. The first 3 program cycles consisted of a 30 s denaturing step at 95 ºC, an annealing step at 5 ºC below the lowest annealing temperature for the overlapping region of each primer pair for 30 s, and an elongation step of 1 min/kb at 72 ºC. The remainder of the program consisted of a 30 s denaturing step at 95 ºC, an annealing step at an annealing step with temperature 5 ºC below the lowest annealing temperature (for the entire primer) of each primer pair for 30 s, and an elongation step of 1 min/kb at 72 ºC. This was repeated 27 times, with each cycle ending in a 10 min extension at 72 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

2.2.3.5 PCR Product Purification

A typical PCR purification consisted of agarose gel electrophoresis employing an 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. PCR products were diluted to 83% with 6X DNA loading dye (0.25% xylene cyanol, 0.25% bromophenol blue, 15% Ficoll 400) and electrophoresed at 100 V for 30 min. After electrophoresis, DNA bands of the correct base pair length, as estimated from comparison with DNA ladder, were excised from the gel with a clean scalpel. DNA was purified from agarose slices using the Qiagen gel extraction kit, according to the manufacturer's protocols with the following

modifications. Instead of heating at 55 ºC to melt the agarose, gel slices were heated at 65 ºC until the gel was fully dissolved. DNA elution buffer EB was also heated to 65 ºC prior to DNA elution, and the volume used for elution was 30 µL.

2.2.3.6 DNA Ligations and Plasmid Transformation

PCR products for the kijanimicin biosynthetic protein constructs were digested with the appropriate restriction enzymes, generally *Nde*I and *Hind*III at 37 ºC overnight. A typical reaction mixture contained 30 µL of purified PCR product, 1-2 µL of each enzyme, 0.4 µL of 100X BSA, 4 µL of 10X NEB buffer 2 and sterile water to a final volume of 40 µL. Digested products were cleaned with the Qiagen PCR purification kit, following the manufacturer's protocols, and ligated to the corresponding sites in the pET28 a or b+ plasmid. The pET28 a or b+ plasmid was prepared by digesting with *Nde*I and *Hind*III at 37 ºC overnight followed by treatment for 1-2 h at 37 ºC with CIP in NEB buffer #2. A typical reaction mixture contained 30 µL of purified vector, 2 µL of each enzyme, 0.4 µL of 100X BSA, 4 µL of 10X NEB buffer 2 and sterile water to a final volume of 40 µL. After overnight digestion, pET28a or b+ was treated for 1-2 h at 37 ºC with 0.2 µL of CIP in NEB buffer #2. The digested plasmid was electrophoresed using a 0.8% TAE-agarose gel, followed by extraction using the Qiagen Gel Extraction kit with modified protocols as previously detailed. In some cases, plasmids were extracted using the Qiagen gel extraction kit, following the instructions for PCR purification.

A Nanodrop instrument was used to quantify the plasmid and the insert DNA solutions. A typical ligation mixture included 90 fm ends of gel purified PCR product, 30 fm ends of pET28a or b+, 2 µL of 5X T4 ligase buffer, 3.5 µL of sterile water, and 0.5 µL of T4 ligase to a final volume of 10 µL. The ligation mixture was incubated at room

temperature for 1-2 h, or overnight at 4 ºC. An aliquot of the ligation mixture (5 µL) was introduced to competent *E. coli* DH5α cells (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, and incubated on ice for 2 min. LB broth (500 µL) was added to the transformed cells and incubated at 37 ºC with 200 RPM shaking for 45 min, and plated on LB-agar plates supplemented with 50 µg/mL kanamycin. Positive transformants were selected by incubation overnight at 37 ºC.

2.2.3.7 DNA Extraction by Alkaline Lysis

Colonies appearing after overnight incubation were screened for the presence of plasmid containing PCR product using the method of alkaline lysis.[277] Briefly, 5 mL cultures of each positive transformant were grown overnight at 37 ºC in LB media with 50 µg/mL kanamycin. Cultures were centrifuged at 12,000 *g* for 30 s, the supernatant was removed and the cells were resuspended in 100 µL of cold Resuspension solution (50 mM glucose, 25 mM Tris•HCl (pH 8.0), 10 mM EDTA (pH 8.0), 100 µg/mL RNase). Cells were disrupted with the addition of 200 µL fresh Lysis solution (200 mM NaOH, 1% SDS), and inverted 5 times to mix. Lysis reactions were quenched with the addition of 150 µL of Neutralization solution (3 M potassium acetate, 11.5% acetic acid). Reactions were mixed to ensure neutralization and incubated at 4 ºC for 5 min, followed by centrifugation for 5 min at 12,000 *g* to precipitate cellular debris. Supernatants containing plasmid DNA were transferred to fresh tubes and the DNA was precipitated by the addition of two volumes of 100% ethanol and 1/10 volume of 3 M sodium acetate, pH 5.2. Solutions were mixed to ensure homogeneity and centrifuged at 12,000 *g* for 15 min at 4 ºC. The supernatants were removed by aspiration, taking care not to disturb the

DNA pellet. DNA pellets were washed with 70% ethanol and centrifuged again at 12,000 *g* for 5 min. The final ethanol wash was removed and solutions were dried using a centrifugal vacuum apparatus for 10 min at room temperature to remove any remaining ethanol. Pellets containing plasmid DNA were resuspended in 30 µL of 50 mM Tris•HCl (pH 8.5).

2.2.3.8 DNA Analytical Restriction Digestion and Fragment Analysis

Plasmid DNA obtained from this preparation was subjected to enzymatic digestion with *Nde*I and *Hind*III. A typical reaction mixture contained 15 µL of DNA, 2 µL of 10X NEB reaction buffer #2, 0.1 µL of each enzyme (*Nde*I and *Hind*III), 0.2 µL of 100X BSA and 2.6 µL of sterile water. The reactions were incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis. Samples were diluted 5X in DNA loading dye and electrophoresed for 30 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide (gel concentration).

Electrophoretic results as visualized on agarose gels were used to confirm the presence of the DNA insert of the appropriate length. Colonies determined to contain the vector and successfully ligated PCR product were grown overnight at 37 ºC with 200 RPM shaking in LB media supplemented with 50 µg/mL kanamycin. After overnight incubation, bacteria were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the bacterial pellet using the Qiagen miniprep kit, following the manufacturer's protocols with the following adjustments to the protocol: the optional 500 µL PB wash was always performed to remove additional contaminants, and elution buffer EB was heated to 65 ºC before DNA elution. The purified plasmid was sequenced to confirm the identity of the PCR product.

PHRED sequences obtained from the DNA facility were aligned against the gene sequences previously identified for the target genes though gene cluster sequencing. Both forward and reverse sequences were compared to the originally obtained gene sequences to determine that the sequences were correct with no mutations.

2.2.3.9 SDS-PAGE

The relative molecular mass and purity of enzyme samples were determined using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli.[273] Proteins were electrophoresed using a discontinuous buffer system 25 mM Tris•HCl, 192 mM glycine and 0.1% SDS (pH 8.3).[273] Gels were 75 mm, and 12% acrylamide was used for the separatory gel, while the stacking gel was 4% acrylamide. Protein samples were diluted 2X with SDS-Loading dye (62.5 mM Tris•HCl buffer pH 6.8, containing 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.0025% bromophenol blue, and heated to 100 ºC for 5 min followed by a brief centrifugation step. The protein gels were electrophoresed using a setting of 30 mA per gel, allowing the process to continue until the loading dye ran off the bottom of the gel. Gels were removed from glass plates, and stacking gel removed. A corner of each gel was also cut to preserve orientation of the gel samples. Gels were stained with Coomassie blue (2.5 g/L of Coomassie Brilliant Blue G-250 in acetic acid:water:methanol (1:4:5) by volume) until gels were dark blue.[278] Gels were removed from staining solution, rinsed in ddH_2O, and introduced to destaining solution (ethanol:acetic acid:water in a ratio of 4:5:41 by volume).[278] Protein concentrations were determined according to Bradford using BSA as the standard, and the dye reagent from BioRad.[272]

2.2.3.10 Western Blot Procedure

Western blotting was used to detect the presence of polyhistidine tags on heterologously expressed and purified proteins.[279] SDS-PAGE of elution fractions from Ni-NTA chromatography procedures was performed in duplicate for each protein. The gels were processed as detailed previously up to the staining step, at which point they were assembled into the transfer cassettes. The gels were transblotted onto the nitrocellulose membrane at 100 V for 1 h, using a running buffer of 25 mM Tris base, 192 mM glycine, 20% v/v methanol (pH 8.3). Cassette assembly and electrophoresis were carried out according to the directions provided by BioRad. After transblotting electrophoresis, the membrane was carefully removed from the transfer cassette and the presence of the prestained marker on the blot was visually assessed to ensure transfer was complete. The membrane was transferred to a small plastic box and immersed in 5% non-fat dry milk in Tris buffered saline and Tween (TBST) (20 mM Tris•HCl, 150 mM NaCl, 0.5% v/v Tween® 20, pH 7.5). The membrane was blocked in this solution for 1 h, with agitation, overnight at 4 ºC or at room temperature for 1 h. After which, the solution was removed and the membrane was rinsed thrice in TBST to remove excess blocking solution. The membrane was immersed in primary antibody solution (15 mL TBST, 1:30,000 dilution of monoclonal anti-polyhistidine clone His-1, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was removed from the membrane and the membrane was washed thrice with TBST to remove any excess primary antibody solution remaining on the membrane. The membrane was incubated in secondary antibody solution (15 mL TBST, 1:30,000 dilution of anti-mouse IgG [Fc specific] alkaline phosphatase conjugate, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was removed from the membrane and

the membrane was washed thrice with TBST to remove any excess secondary antibody solution remaining on the membrane. To visualize proteins, the membrane was immersed in 5 mL color development solution (100 mM Tris•HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) with the addition of 66 μL NBT and 33 μL BCIP. The membrane was incubated in the dark for 2-5 min or until purple bands appeared on the blot. Development was stopped by rinsing the membrane with ddH_2O. Membranes were dried and scanned within 24 h.

2.2.4 GENERATION OF TDP-L-DIGITOXOSE BIOSYNTHETIC PROTEINS

2.2.4.1 TTP3 PROTEIN MIXTURE

The genes encoding thymidine kinase (TK), thymidylate kinase (TMK), and nucleotide diphosphate kinase (NDK) were amplified from the genomic DNA of *E. coli* strain HMS174 and cloned into the pET28b+ vector in tandem such that each gene is associated with its own ribosome binding site and His_6-tag, by Dr. Yung-nan Liu. The PCR products of the TK, TMK and NDK genes, after digestion with the appropriate restriction enzymes, *Nde*I/*EcoR*I, *EcoR*I/*Hind*III or *Hind*III/*Xho*I, respectively were ligated in-frame and downstream from the His-tag encoding sequence to the *Nde*I/*Xho*I digested pET28b(+) vector. The resulting plasmid pET/TK/TMK/NDK was used to transform the *E. coli* cloning host DH5α, and kanamycin resistant clones were characterized by restriction mapping. The plasmid DNA isolated from the positive clones was used to transform *E. coli* BL21(DE3).

STEP 1: GROWTH OF E. COLI CELLS. An overnight culture of *E. coli* BL21(DE3)-pET/TK/TMK/NDK was grown in LB media supplemented with kanamycin (50 μg/mL)

at 37 °C. 2 mL of overnight culture was used to inoculate 1 L of LB supplemented with kanamycin (50 μg/mL). Typically, 6 L of culture was incubated at 37 °C until the OD_{600} reached 0.4-0.6, followed by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM, and the temperature was reduced to 25 ºC. The culture was allowed to grow for an additional 18 h at 25 °C, at which time the cells were harvested by centrifugation at 4,500 *g* for 15 min at 4 °C, and stored at -80 °C.

STEP 2: CRUDE EXTRACT PREPARATION. Cells from 6 L of culture were resuspended in 100 mL of Lysis buffer I (50 mM sodium phosphate, 150 mM NaCl, 10 mM imidazole, pH 8.0), and disrupted with 5 × 1 min sonication bursts, with 1 min cooling intervals between bursts. Cellular debris was removed by centrifugation at 15,000 *g* for 25 min, and the soluble supernatant was removed and retained.

STEP 3: NI-NTA CHROMATOGRAPHY. To the crude extract, 4 mL of prepared Ni-NTA resin was added and the resulting mixture was incubated for 1 h with gentle shaking at 4 °C. The resin slurry was loaded into an empty column (2 × 25 cm). The column was washed with 100 mL of Wash buffer I (50 mM sodium phosphate, 150 mM NaCl, 20 mM imidazole, pH 8.0), and the TK, TMK and NDK protein mixture was eluted from the column using 12 mL of Elution buffer I (50 mM sodium phosphate, 150 mM NaCl, 250 mM imidazole, pH 8.0). The desired protein fractions were pooled and desalted by dialyzing against 1 L of Dialysis buffer I (50 mM potassium phosphate buffer (pH 7.5) 15% glycerol) with 3 changes.

2.2.4.2 RFBA- (α -D-GLUCOSE-1-PHOSPHATE THYMIDYLYLTRANSFERASE)

The α-D-glucose-1-phosphate thymidylyltransferase, RfbA was prepared in the following manner. The *rfbA* gene was amplified from *Salmonella enterica serovar*

Typhimurium LT2 genomic DNA with an upstream ribosome binding site (RBS)/translational spacer element (TSE) using PCR by Dr. Yung-nan Liu. The PCR-amplified gene was purified, digested with *Bam*HI and *Pst*I, and ligated into *Bam*HI/*Pst*I digested pUC18. A $(His)_5$-tag was added to the *C*-terminus of RfbA by PCR amplification of the *rfbA* insert of plasmid rfbA/pUC18 and cloning back into the same vector. The rfbA-$(His)_5$/pUC18 plasmid was used to transform the BL21(DE3) cells for protein expression.

STEP 1: GROWTH OF E. COLI CELLS. The resulting construct was used to transform *E. coli* BL-21(DE3). Cultures was grown in LB media in the presence of 100 μg/mL ampicillin overnight. Overnight culture (1 mL) was used to inoculate each L of fresh LB broth. The cultures were subsequently incubated for 20 h at 37 ºC with 250 RPM shaking. Cells were harvested by centrifugation for 15 min at 4,500 *g* at 4 ˚C and stored overnight at -80 ˚C.

STEP 2: CRUDE EXTRACT PREPARATION. The protein RfbA is expressed as an N-terminal His_6 tagged protein and purified by affinity chromatography with Ni-NTA resin. The cells were resuspended in 2 mL Lysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole with 10% glycerol pH 8.0) per gram cell wet weight. Cells were disrupted by 12 × 30 s sonication pulses with 1 min cooling periods between the pulses and crude protein extracts were obtained by centrifugation at 10,000 *g,* 4 ºC for 20 min.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was further purified by Ni^{2+} affinity chromatography following manufacturer's instructions with 10% glycerol in all buffers. Briefly, 4 mL of resin was prepared by washing with Lysis buffer II and allowed to bind with soluble protein at 4 ºC for 1 h, then loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 column volumes (CV) of Lysis buffer followed by two washes with 10 CV of Wash buffer II (50 mM NaH_2PO_4, 300 mM NaCl,

20 mM imidazole, 10% glycerol, pH 8.0). RfbA protein was eluted with 12 mL of Elution buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 15% glycerol, pH 8.0), which was collected in 1.5 mL fractions, with RfbA most heavily concentrated in fractions E3-E5. These fractions were collected and dialyzed against 3 L of Dialysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 15% glycerol, pH 8.0). The dialyzed protein solution was concentrated using the Amicon filtration system. Protein was flash frozen in liquid nitrogen and stored at -80 ºC until use.

2.2.4.3 RFBB (TDP-GLUCOSE 4,6-DEHYDRASE)

RfbB used in the synthesis was prepared in the following manner. The *rfbB* gene was amplified from *Salmonella enterica serovar Typhimurium* LT2 genomic DNA using PCR by Dr. Yung-nan Liu. The PCR-amplified gene was purified, digested with *Eco*RI and *Bam*HI, and ligated into *Eco*RI/*Bam*HI digested pUC18. An *N*-terminal His_6 tag was added to the construct in a manner similar to that previously described for the recombinant RfbA construct.

STEP 1: GROWTH OF E. COLI CELLS. The resulting construct was used to transform *E. coli* BL21(DE3) and grown in LB broth in the presence of 100 µg/mL ampicillin overnight at 37 ºC. Overnight culture (1 mL) was used to inoculate each liter of fresh LB broth. The cultures were subsequently incubated for 20 h at 37 ºC with 250 RPM shaking. Cells were harvested by centrifugation for 15 min at 4,500 *g* at 4 ˚C and stored overnight at -80 ˚C.

STEP 2: CRUDE EXTRACT PREPARATION. The protein RfbB is expressed as an N-terminal His_6 tagged construct and purified by affinity chromatography using Ni-NTA resin. The cells were resuspended in 2 mL Lysis buffer with 10% glycerol pH 8.0 per

gram cell wet weight. Cells were disrupted by sonication 12 × 30 s pulses with 1 min cooling periods and crude protein extracts were obtained by centrifugation for 20 min at 4 ºC at 10,000 *g*.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was further purified by Ni^{2+} affinity chromatography following manufacturer's instructions with 10% glycerol in all buffers. Briefly, 6 mL of resin was prepared by washing with Lysis buffer II and allowed to bind with soluble protein at 4 ºC for 1 h, then loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 CV of Lysis buffer followed by two washes with 10 CV of Wash buffer II. RfbB protein was eluted with 12 mL of Elution buffer II, which was collected in 1.5 mL fractions, with RfbB most heavily concentrated in fractions E2-E8. These fractions were collected and dialyzed against 3 L of Dialysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 15% glycerol, pH 8.0). The dialyzed protein solution was concentrated using the Amicon filtration system. Protein was flash frozen in liquid nitrogen and stored at -80 ºC until use.

2.2.4.4 TYLX3 – (2,3-DEHYDRATASE)

Previous Liu group members have used the KijB1 homolog TylX3 to generate the TDP-α-2,6-dideoxy-D-glycero-4-hexulose. Briefly, Dr. Huwai (Raymond) Chen created the expression construct by amplifying the *tylX3* gene by PCR from template DNA obtained from *S. fradiae*. The PCR product was purified, digested and ligated into the pET28b+ vector to generate an *N*-terminally-His_6 tagged gene construct. The ligation mixture was used to transform *E. coli* DH5α cells, and colonies positive for the construct pHC45 were isolated and grown overnight at 37 ºC with 250 RPM shaking in LB broth supplemented with 50 µg/mL kanamycin.

STEP 1: GROWTH OF E. COLI CELLS. The pHC45 plasmid was purified using the Qiagen miniprep kit following the manufacturer's protocols and used to transform *E. coli* BL-21(DE3). After overnight selection on LB-agar supplemented with 50 µg/mL kanamycin, colonies that grew were used to inoculate 25 mL of LB broth supplemented with kanamycin at 50 µg/mL, and grown overnight at 37 ºC with 250 RPM shaking. The overnight culture was used to inoculate 6 L of fresh LB media (4 mL culture per L broth), supplemented with 50 µg/mL kanamycin. The cultures were grown at 37 ºC with 250 RPM shaking until the OD_{600} of the culture reached 0.3, at which point the temperature was reduced to 24 ºC. The cultures grown for an additional h or until the OD_{600} reached 0.4. The cultures were induced with IPTG to a final concentration of 50 µM, and allowed to grow for an additional 15 h at 24 ºC. The cells were harvested by centrifugation at 4,500 *g* for 20 min at 4 ºC and stored at –80 ºC until protein purification.

STEP 2: CRUDE EXTRACT PREPARATION. The cells were resuspended in 2 mL Lysis buffer III (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 20% glycerol, 5 mM β-mercaptoethanol, pH 8.0) per gram of cells (wet weight). The resuspended cells were disrupted by sonication using 5 × 1 min pulses, with 1 min cooling intervals between bursts. After sonication, the cellular debris was removed by centrifugation at 35,000 *g* for 20 min. The supernatant was decanted and retained for further purification.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was further purified by Ni^{2+} affinity chromatography following slightly adjusted manufacturer's instructions, with 20% glycerol and 5 mM β-mercaptoethanol in all buffers. Briefly, 5 mL of resin was prepared by washing with Lysis buffer III and allowed to bind with soluble protein at 4 ºC for 2 h with gentle shaking. The mixture was centrifuged at 1,000 *g* and the supernatant containing unbound protein was decanted from the pelleted resin. The resulting slurry was mixed well and loaded onto a 2 × 25 cm column. The column was packed by gravity

flow and washed with 1 CV of Lysis buffer III followed by two washes with 5 CV each of Wash buffer IIIa (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 20% glycerol, 5 mM β-mercaptoethanol, pH 8.0) and IIIb (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM imidazole, 20% glycerol, 5 mM β-mercaptoethanol, pH 8.0). The TylX3 protein was eluted with 20 mL of Elution buffer III (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 20% glycerol, 5 mM β-mercaptoethanol, pH 8.0) and collected in 1.5 mL fractions. The pooled fractions were dialyzed against 4 L of Dialysis buffer III (50 mM Tris, 20% glycerol, pH 7.5 at 25 ºC). The dialyzed protein solution was concentrated using the Amicon stirred cell filtration system. Protein was flash frozen in liquid nitrogen and stored at -80 ºC until use.

2.2.4.5 KijD10 (3-Ketoreductase)

This gene was amplified by PCR from a cosmid library constructed from *A. kijaniata* using the primers (upstream 5′-TAGGGGAGTT*CATATG*GAGAATCCGGC-GAA-3′) and (downstream 5′-GGCTCCTGAA*AAGCTT*CTAAACCCGGGTCA-3′). The PCR fragment was subsequently purified by agarose gel electrophoresis followed by Qiagen gel extraction kit and ligated into the vector pANT841. This construct was subsequently digested with *Nde*I and *Hind*III and the fragment ligated into corresponding sites in the vector pET28b to create a protein expression construct with an *N*-terminal His_6 tag. The construct was isolated from *E. coli* DH5α using the Qiagen miniprep kit, and submitted for DNA sequencing to ensure there were no mutations from the PCR process.

Step 1: Growth of E. coli Cells. This construct was used to transform *E. coli* BL21(DE3) pLysS cells which were subsequently grown in 10 mL LB media

supplemented with 50 μg/mL kanamycin at 37 ºC with 250 RPM shaking. The cultures were transferred to 6 L of fresh LB media and incubated at 37 ºC with 250 RPM shaking until the OD_{600} reached 0.4. The temperature was reduced to 14 ºC and protein expression was induced with 1 mM IPTG. Cells were grown for 18 h after induction and subsequently harvested by centrifugation for 15 min at 4,500 *g* and frozen overnight at -80 ºC.

STEP 2: CRUDE EXTRACT PREPARATION. The cells were resuspended in 4 mL of Lysis buffer II, per gram cell, wet weight. Cells were disrupted by sonication using 24 × 10 s pulses, with 20 s cooling intervals between bursts. Crude protein extracts were obtained by centrifugation at 10,000 *g* at 4 ºC for 20 min.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was further purified by Ni^{2+} affinity chromatography following manufacturer's protocols with 10% glycerol in all buffers. Briefly, 10 mL of resin was prepared by washing with Lysis buffer II and allowed to bind with soluble protein at 4 ºC overnight, and loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 CV of Wash buffer II. KijD10 protein was eluted with 15 mL of Elution buffer II and collected in 1.8 mL fractions, with KijD10 most heavily concentrated in early fractions E1-E4. These fractions were combined and concentrated by filtration using a YM-10 and dialyzed against Dialysis Buffer IV (50 mM Tris•HCl, 2 mM EDTA, 1 mM DTT with 10% glycerol, pH 7.5). Protein was flash frozen in liquid N_2 and stored at -80 ºC until use.

2.2.4.6 KIJD11 (5-EPIMERASE)

This gene was amplified by PCR from a cosmid library constructed from *A. kijaniata* using the primers (upstream 5′-AGGACATG*CATATG*AGCGGTATCGA-

GATCG-3′) and (downstream 5′-GGTGTCGCGG*AAGCTT*CCGCCCGGCTCAGA-3′). The PCR fragment was purified by gel extraction (Qiagen) and ligated into the vector pANT841. This construct was subsequently digested with *Nde*I and *Hind*III and the fragment ligated into pET28a to create a protein expression construct with an N-terminal His_6 tag. The construct was isolated from *E. coli* DH5α using the Qiagen miniprep kit, and submitted for DNA sequencing to confirm the correct gene sequence.

STEP 1: GROWTH OF E. COLI CELLS. This construct was use to transform *E. coli* Bl21(DE3) pLysS cells which were subsequently grown in 10 mL LB media supplemented with 50 µg/mL kanamycin and 35 µg/mL chloramphenicol at 37 ºC with 250 RPM shaking. The cultures were transferred to 6 L of fresh LB media supplemented with both 50 µg/mL kanamycin and 35 µg/mL chloramphenicol and incubated at 37 ºC with 250 RPM shaking until the OD_{600} reached 0.5-0.6. The temperature was reduced to 14 ºC and protein expression was induced with the addition of 1 mM IPTG. Cells were incubated for an additional 20 h. Cells were harvested by centrifugation for 15 min at 4,500 *g* and frozen overnight at -80 ºC.

STEP 2: CRUDE EXTRACT PREPARATION. The cells were resuspended in Lysis buffer II, using 2 mL buffer per gram cell wet weight, to which 1 mg/mL lysozyme was added. This mixture was allowed to incubate on ice for 1 h, stirring every 15 min. Cells were disrupted by sonication using 5 × 20 s pulses, with 40 s cooling intervals between bursts. Crude protein extracts were obtained by centrifugation at 10,000 *g* at 4 ºC for 20 min.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was purified by Ni^{2+} affinity chromatography following manufacturer's instructions with 10% glycerol in all buffers. Briefly, 5 mL of resin was prepared by washing with Lysis buffer II and allowed to bind with soluble protein at 4 ºC for 1 h, then loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 CV of Lysis buffer II followed by two

washes with 10 CV of Wash buffer II. KijD11 protein was eluted with 15 mL of Elution buffer I and collected in 1.5 mL fractions, with KijD11 most heavily concentrated in fractions E1-E5. These fractions were collected and concentrated by filtration using a YM-10 and dialyzed against Dialysis Buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 15% Glycerol pH 8.0) Protein was flash frozen in liquid N_2 and stored at -80 ºC until use.

2.2.4.7 KijC2 (4-Ketoreductase)

This gene was amplified by PCR from a cosmid library constructed from *A. kijaniata* using the primers (upstream 5′-AGGACATG*CATATG*AGCGGTATCGA-GATCG-3′) and (downstream 5′-GGTGTCGCGG*AAGCTT*CCGCCCGGCTCAGA-3′). The PCR product was subsequently purified by using the Qiagen gel extraction kit, digested with *Nde*I and *Hind*III and ligated into the corresponding sites of the vector pET28a to create a protein expression construct with an *N*-terminal His_6 tag. The construct was isolated from *E. coli* DH5α using the Qiagen miniprep kit, and submitted for DNA sequencing to confirm the correct gene sequence.

Step 1: Growth of E. coli Cells. This construct was used to transform *E. coli* BL21 Rosetta II DE3 (Novagen) cells by heat shock and grown in 10 mL LB media supplemented with 50 µg/mL kanamycin and 35 µg/mL chloramphenicol at 30 ºC with 250 RPM shaking. After overnight incubation, 3 mL of culture (per L) was transferred to 6 L of fresh LB media supplemented with 50 µg/mL kanamycin and 35 µg/mL chloramphenicol, and incubated at 30 ºC with 250 RPM shaking for 24 h. Cells were harvested by centrifugation for 15 min at 4,500 *g* and frozen overnight at -80 ºC.

Step 2: Crude Extract Preparation. The cells were resuspended in 2 mL Lysis buffer IV (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 1 mM DTT with 10%

glycerol, pH 8.0) per gram cell wet weight, to which 1 mg/mL lysozyme was added. This mixture was allowed to incubate on ice for 1 h, stirring every 15 min. Cells were disrupted by sonication using 8 × 45 s pulses, with 1 min cooling intervals between bursts. Crude protein extracts were obtained by centrifugation at 10,000 *g* at 4 ºC for 20 min.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was purified by Ni^{2+} affinity chromatography following manufacturer's instructions with 10% glycerol in Lysis and Wash buffers, and 15% glycerol in Elution buffer. 1 mM dithiothreitol (DTT) was included in all buffers. Briefly, 6 mL of resin was prepared by washing with Lysis buffer IV (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 10% glycerol, pH 8.0) and allowed to bind with soluble protein at 4 ºC for 2 h, then loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 CV of Lysis buffer followed by two washes with 10 CV of Wash buffer IV (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 1 mM DTT, 10% glycerol, pH 8.0). KijC2 protein was eluted with 15 mL of Elution buffer IV (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 1 mM DTT, 20% glycerol, pH 8.0) and collected in 1.5 mL fractions, with KijC2 most heavily concentrated in fractions E4-E7. These fractions were collected and concentrated by filtration using a YM-10 and dialyzed against Dialysis Buffer IV (50 mM NaH_2PO_4, 300 mM NaCl, 20% glycerol, 1 mM DTT, pH 8.0). Protein was flash frozen in liquid N_2 and stored at -80 ºC until use.

2.2.5 TDP-L-DIGITOXOSE BIOSYNTHESIS

2.2.5.1 Enzymatic Synthesis of TDP-D-glucose (2)

TDP-α-D-glucose was prepared from thymidine and glucose-1-phosphate (**1**) in a two-stage, one-pot reaction.[280] A reaction mixture containing 76.2 mM phosphoenolpyruvate (PEP), 24 mM thymidine, 1.6 mM ATP, 27 mM $MgCl_2$, 25 μM thymidine kinase (TK), 25 μM thymidylate kinase (TMK), 25 μM nucleoside diphosphate kinase (NDK), and 1000 units of rabbit muscle pyruvate kinase (PK) in 17 mL of 45 mM Tris•HCl buffer, pH 7.5, was incubated at 37 °C for 4 h, generating thymidine triphosphate (TTP). The enzymes were removed by filtration through an Amicon YM-10 membrane at 4 ºC, and glucose-1-phospate (**4**), $MgCl_2$, and α-D-glucose-1-phosphate thymidylyltransferase (RfbA) from *S. enterica* LT2 were added to the filtrate to give final concentrations of 28 mM, 50 mM, and 36 μM, respectively. The mixture was incubated for 16 h at 37 °C, centrifuged at 5,000 *g* for 10 min at 4 ºC to remove precipitate, and filtered through an Amicon YM-10 membrane at 4 ºC to remove enzymes. The crude product (**5**, with a theoretical yield of 228 mg), was stored at 4 °C.

2.2.5.2 Purification of TDP-D-glucose (5) by Bio-Gel P2 Chromatography

The enzyme-free filtrate from the previous step was loaded onto a Bio-gel P2 column (25 mm × 100 cm) pre-washed with water, and run at a flow rate of 12 mL/h with water as the eluent, collecting 8 mL fractions. Fractions displaying UV absorption at 267 nm were lyophilized, and the identity and purity of the compounds in each fraction analyzed by HPLC using the analytical program and conditions detailed below. TDP-D-glucose (**5**) containing fractions (total weight 170 mg) varied in purity from 25-70%, were pooled according to their purities.

2.2.5.3 Enzymatic Conversion of TDP-D-glucose (5) to TDP-4-keto-6-deoxy-D-glucose (6)

TDP-D-glucose (**5**, 25 mg) obtained from the previous step was dissolved in 50 mM Tris buffer, pH 7.5 to give a final concentration of 23 mM. This solution was incubated with the 4,6-dehydrase RfbB from *S. enterica* LT2 (18 μM) at 37 °C for 2 h, or for 3-4 hour at 30 ºC, after which RfbB was removed by filtration through an Amicon YM-10 membrane at 4 ºC. A total of 23.75 mg of TDP-4-keto-6-deoxy-D-glucose (**6**), which was greater than 90% pure, was obtained. The concentration of **6** was determined spectrophotometrically at 267 nm, where the molar extinction coefficient (ε) of **6** is 9600 $M^{-1}cm^{-1}$.

2.2.5.4 Enzymatic Conversion of TDP-4-keto-6-deoxy-D-glucose (6) to TDP-4-keto-2,6-dideoxy-D-glucose (8)

TDP-4-keto-6-deoxy-D-glucose (**6**) 23.75 mg, obtained from the previous step was collected and added to a mixture containing 160 nM TylX3 (KijB1 substitute), 360 nM KijD10, 1.2 M substrate equivalents (~51 mg) NADPH, 10% glycerol in 50 mM Tris buffer, pH 7.5 for a final reaction volume of 3 mL. This reaction was allowed to incubate at room temperature (25 ºC) for 8-12 h. The progress of the reaction was monitored by HPLC using the analytical program and conditions detailed below. When the reaction was observed to have come to completion, the enzymes were removed by filtration through an Amicon YM-10 membrane at 4 ºC. Approximately 20 mg of TDP-4-keto-2,6-dideoxy-D-glucose (**8**) was obtained.

2.2.5.5 Purification of TDP-4-keto-2,6-deoxy-D-glucose (8) by Bio-Gel (P2) Chromatography

The enzyme-free filtrate from the previous step was loaded onto a Bio-Gel P2 column (25 mm × 100 cm) pre-washed with water, and run at a flow rate of 12 mL/h with 25 mM NH_4HCO_3 at ambient pH as the eluent, collecting 8 mL fractions. Fractions observed to have UV absorption at 267 nm were lyophilized to near dryness, and the identity and purity of the compounds in each fraction assessed by HPLC using the analytical program and conditions detailed below. TDP-4-keto-2,6-dideoxy-D-glucose (**8**) containing fractions generally eluted after 22-24 h with the compound of interest in 2-3 fractions only. Generally, the two later fractions were of very high purity (> 90%) and one early fraction was often contaminated with $NADP^+$, a byproduct of the reduction reaction. Overall recovery of product (**8**) varied from 10 to 15 mg. The two high-purity, high-concentration fractions were combined and lyophilized to near dryness (generally to < 1 mL total volume). NMR spectra: ^{1}H NMR (300 MHz, D_2O): (a mixture of hydrate and keto forms): δ 1.10 (3H, d, *J* = 6.4, 5-Me of the hydrate form), 1.18 (3H, d, *J* = 6.4, 5-Me of the keto form), 1.72-1.78 (1H, m, 2-H_{ax}), 1.82 (3H, s, 5″-Me), 2.05 (1H, m, 2-H_{eq}), 2.20-2.30 (2H, m, 2′-H), 3.91 (1H, dd, *J* = 12.9, 5.9, 3-H), 3.94 (1H, q, *J* = 6.4, 5-H), 4.06 (3H, m, 4′-H, 5′-H), 4.51 (1H, m, 3′-H), 5.50 (1H, dd *J* = 6.6, 4.0, 1-H), 6.24 (1-H, t, *J* = 6.9, 1′-H), 7.64 (1H, s, 6″-H). ^{31}P NMR (121 MHz, D2O): δ -10.6 (d, *J* = 19.7), -12.6 (d, *J* = 19.7). High resolution ESI-MS of **8**, calculated for $C_{16}H_{25}N_2O_{15}P_2$ (M-H)-547.0725, found 547.0691.

2.2.5.6 Enzymatic Conversion of TDP-4-keto-2,6-dideoxy-D-glucose (8) to TDP-L-Digitoxose (10)

The concentration of the product of the previous step, TDP-4-keto-2,6-dideoxy-D-glucose (**8**), was determined spectrophotometrically at 267 nm, where the molar extinction coefficient (ε) of **TDP** is 9600 $M^{-1}cm^{-1}$. Product (**8**) was added to a mixture containing 100 nM KijD11, 10 μM KijC2, 1.4 M substrate equivalents NADH, 10% glycerol and 10 mM $MgCl_2$ in 50 mM Tris buffer, pH 7.5 with the final reaction volume adjusted to bring the concentration of **8** to 1 mM. The mixture was divided into 200 μL aliquots and allowed to incubate at room temperature (25 ºC) for 12-18 h with gentle agitation. The progress of the reaction was monitored by HPLC using the analytical program and conditions detailed below. When the reaction was observed to have come to completion, the enzymes were removed by filtration through an Amicon YM-10 membrane at 4 ºC. Approximately 0.1 mg of product TDP-L-digitoxose (**10**) was obtained for every mg TDP-4-keto-2,6-dideoxy-D-glucose (**8**) substrate used.

2.2.5.7 Purification of TDP-L-Digitoxose (10) using FPLC and Bio-Gel (P2) column chromatography

The enzyme-free filtrate from the previous step was purified using a Fast-Protein Liquid Chromatography (FPLC). All FPLC analysis was performed using a Mono Q 16/10 column with 500 μL of sample for each injection. The sample was eluted by a gradient of water as solvent A and 250 mM NH_4HCO_3 as solvent B where the gradient ran from 0 to 40% B over 5 min, held at 40% for 5 min, running from 40 to 100% B over 20 min, 60 to 100% B over 2 min, 10 min wash at 100% B, 100 to 5% B over 5 min, and reequilibration at 5% B for 15 min. The flow rate was 4 mL/min and fractions were

collected beginning at 50% B. TDP-L-digitoxose eluted at 55-60% B and fractions were immediately flash frozen in liquid nitrogen and stored at -80 ºC.

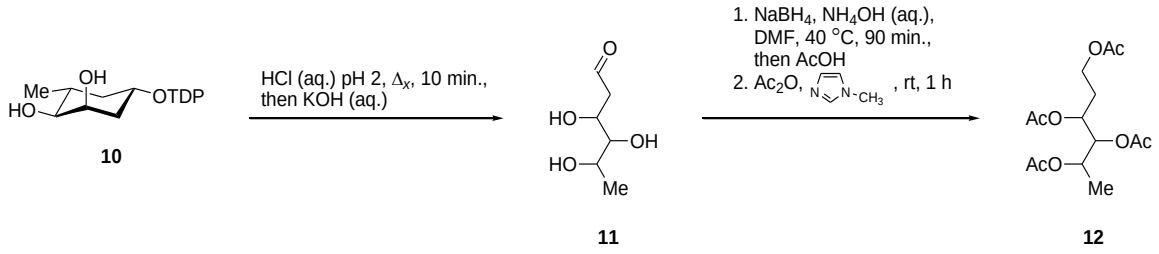
Fractions that appeared to contain the compound of interest (**10**) were analyzed by HPLC using the analytical program detailed below, to verify the presence of TDP-L-digitoxose (**10**). The fractions containing **10** were stored at -80 ºC and thawed gently at 4 ºC. The compound was desalted using a Bio-Gel (P2) column (1 × 12 cm) with the flow rate set at 5 mL/min with water as the eluent, collecting 8 mL fractions. Fractions collected from this column were spectrophotometrically analyzed at 267 nm to detect the presence of TDP. Fractions showing UV absorption at 267 nm were flash frozen in liquid N_2 and stored at -80 ºC until concentration by lyophilization. Samples (20 μL) of each eluate fraction showing UV absorbance at 267 were analyzed by HPLC to determine the identity and purity of the compounds in each fraction. The resulting product was concentrated by lyophilization and re-purified to remove any remaining salt or glycerol by HPLC using a C-18 column (Varian), monitoring at 267 nm, with a flow rate of 1 mL/min, using water as the eluent. Low resolution CI-MS of **10**, calculated for $C_{16}H_{26}N_2O_{14}P_2$ (M + H) 532, found 532. Low resolution ESI of **10**, calculated for $C_{16}H_{26}N_2O_{14}P_2$ (M - 2H) 530, found 265. Attempts to obtain high resolution MS data (FAB, CI, ESI) and ^{1}H NMR spectrum were unsuccessful.

2.2.5.8 TDP-L-Digitoxose Acetylation

Following the procedure reported by Blakeney et al. and previously used in our research group as shown in scheme 2.6, compound **10** (approx. 0.5 mg, 90 μmol, 1 eq) was acidified with concentrated HCl (approx. 200 μL) to pH 2 and was boiled for 10 min. [281, 282] This mixture was neutralized with 12 M potassium hydroxide (KOH) (pH 7 as

judged by pH paper) and lyophilized to dryness. The mixture was redissolved in 0.5 M NH_4OH (approx. 0.6 mL), which was subsequently mixed with 1 mL dry dimethyl formamide (DMF). DMF was dried by pouring over hot 4Å mol sieves. To this was added $NaBH_4$ (21 mg, 555 µmol, 6.2 eq). The reaction was stirred at 40 ºC for 90 min and quenched with 100 µL of glacial acetic acid (at room temp, ~21 ºC). 1-Methyl-imidazole (200 µg, 2 µmol, 0.02 eq) and 6 mL of freshly distilled acetic anhydride were added to the mixture (at room temperature, 21 ºC) and the reaction was allowed to stir at room temperature (21 ºC) for just over 1 h. The reaction was quenched with 5 mL cold methanol. The resulting crude reaction mixture was partitioned between water and chloroform, and the aqueous layer was extracted with chloroform (2 × 30 mL). Combined organic layers were washed with water (2 × 10 mL) and dried (Na_2SO_4), and concentrated. The resulting sample was subjected to GC/MS. Low-resolution GC/EI-MS of **12**: 318 (the parent peak), 231 (C1-C4 and C3-C6 fragments), 245 (C2-C6 fragment).

2.6 Acetylation of TDP-L-digitoxose



2.2.5.9 Characterization of the Degradation Product (2S,3S)-2-methyl-3-hydroxy-4-keto-2,3-dihydropyran (9a) by ^{1}H NMR

To verify the activity of KijD11, a time course assay was conducted and monitored by *in situ* NMR spectroscopy. The protein was prepared by dialyzing a fresh preparation of KijD11 against 6 L (12 × 500 mL) of dialysis buffer IIb (50 mM NaH_2PO_4, 300 mM NaCl, pH 8) without glycerol at 4 ºC, exchanging the buffer every 30 min, and allowing the last exchange to incubate overnight at 4 ºC. After dialysis, the protein was removed from the dialysis tubing and centrifuged for 20 min at 10,000 *g* (4 ºC) to remove any precipitate. The protein was quantified following the method of Bradford.[272] Compound (**8**) TDP-4-keto-2,6-dideoxy-D-glucose of high purity was lyophilized, and redissolved with 100 µL D_2O and an appropriate amount of 50 mM potassium phosphate buffer (pH 7.5) to achieve a final concentration of 10 mM in 600 µL in an NMR tube. The initial spectrum was acquired, and a special program was applied to reduce the water peak. The tube was then removed and 5 µL of KijD11 was added to a final enzyme concentration of 10 nM. The assay was monitored for 3 h with time points taken every 5 min throughout this experiment, using the initial reading (prior to enzyme addition) as time point 0 (see figure 2.9). At the completion of the reaction, a NOSY NMR was also performed. Product (**8**) ^{1}H NMR (500 MHz, D_2O): δ 1.33 (3H, d, J = 6.4, 5-Me), 1.68-1.76 (1H, m, 2-H_{ax}), 1.82 (3H, d, J = 4, 5″-Me), 2.20-2.30 (2H, m, 2′-H), 2.24-2.30 (1H, m, 2-H_{eq}), 3.55-3.60 (1H, m, 5-H), 3.86-3.96 (1H, m, 3-H), 4.06 (3H, m, 4′-H, 5′-H), 5.18 (1H, m, 1-H), (1-H, t, J = 6.8, 1′-H), 7.64 (1H, s, 6″-H). Degradation product (**9a**) ^{1}H NMR (500 MHz, D_2O) δ 1.41 (3H, d, J = 6.9, 2-Me), 4.12 (1H, d, J= 12.4, 3-H), 4.26-4.34 (1H, m, 2-H), 5.37 (1H, d, J = 5.8, 5-H), 7.46 (1H, d, J = 5.8, 6-H).

2.2.5.10 Analytical HPLC Program and Conditions

All HPLC analysis was performed using an analytical Dionex Carbopac PA1 (4 cm × 250 cm) column, equipped with a guard column (4 cm × 50 cm) with 20 μL of sample for each injection. The sample was eluted by a gradient of water as solvent A and 500 mM NH_4OAc (adjusted to pH 7.0 with aqueous NH_3) as solvent B; where the gradient ran from 5 to 20% B over 15 min, then 20 to 60% B over 20 min, 60 to 100% B over 2 min, 3 min wash at 100% B, 100-5% B over 5 min, and re-equilibration at 5% B for 15 min. The flow rate was 1 mL/min, and the detector was set at 267 nm to observe TDP.

2.3 RESULTS AND DISCUSSION

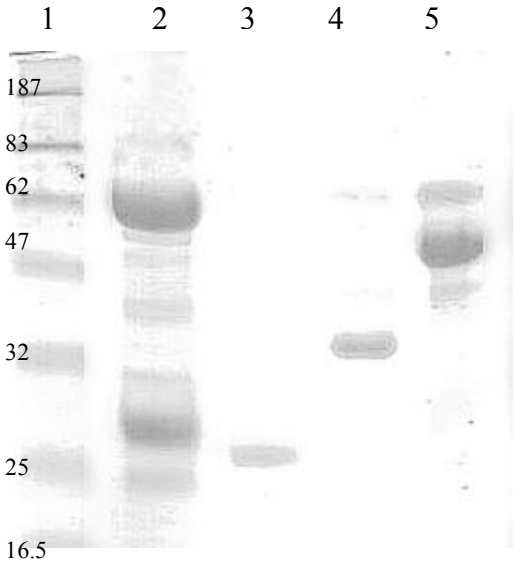
To reconstruct the TDP-L-digitoxose pathway *in vitro*, the enzymes proposed to be involved in the biosynthetic pathway were cloned and heterologously expressed in *E. coli*. The proteins were purified and used in the subsequent biosynthetic assays.

2.3.1 PROTEIN PURIFICATION

A total of 9 enzymes were purified from *E. coli* as His_6-tagged proteins using Ni-NTA chromatography. Native *E. coli* TDP biosynthetic enzymes thymidine kinase (TK), thymidylate kinase (TMK), and nucleotide diphosphate kinase (NDK) were purified together, thus specific purification data for each enzyme is unavailable. After purification, the thymidylyltransferase RfbA protein concentration was determined to be 45.7 mg/mL or 1.404 mM. RfbA has a molar extinction coefficient (ε) of 30,820 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 32.6 kDa. The approximate amount of RfbA purified by this protein procedure was 140 mg; or 23 mg of RfbA per L of LB Broth. The protein

concentration of purifed 4,6-dehydratase RfbB concentration was determined to be 80 mg/mL or 1.93 mM. RfbB has a molar extinction coefficient (ε) of 65,790 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 41.6 kDa. The approximate amount of RfbB purified by this protein procedure was 480 mg; or 80 mg of RfbB per L of LB Broth. The KijB1 homolog, TylX3 was originally purified by Dr. H. Chen. The protein concentration following Dr. Chen's method was determined to be 4 mg/mL or 76 μM. TylX3 has a molar extinction coefficient (ε) of 60,650 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 52.4 kDa. The approximate amount of TylX3 purified by this protein procedure was 60 mg; or 10 mg of TylX3 per L of LB Broth. The putative 3-ketoreductase, KijD10 protein concentration was determined to be 5.4 mg/mL or 145 μM using our purification methods. KijD10 has a molar extinction coefficient (ε) of 38,810 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 37.1 kDa. The approximate amount of KijD10 purified by this protein prep was 54 mg; or 9 mg of KijD10 per L of LB Broth. The putative 5-epimerase, KijD11, protein concentration was determined to be 35 mg/mL or 1.6 mM. KijD11 has a molar extinction coefficient (ε) of 19,420 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 19.4 kDa. The approximate amount of KijD11 purified by this protein prep was 70 mg; or 12 mg of KijD11 per L of LB Broth. The final protein, KijC2 was purified to a concentration of 28.6 mg/mL or 820 μM. The putative 4-ketoreductase KijC2 has a molar extinction coefficient (ε) of 32,530 $M^{-1}cm^{-1}$ at A_{280} and a molecular mass of 34.9 kDa, but is observed as a dimer of ~ 70 kDa. The approximate amount of KijC2 purified by this protein prep was 42 mg; or 7 mg of KijC2 per L of LB Broth. Proteins KijB1, KijD10, KijD11 and KijC2 are shown below in figure 2.5.

2.5 TDP-L-digitoxose Biosynthetic Proteins



1	Molecular Weight Marker
2	KijC2
3	KijD11
4	KijD10
5	KijB1

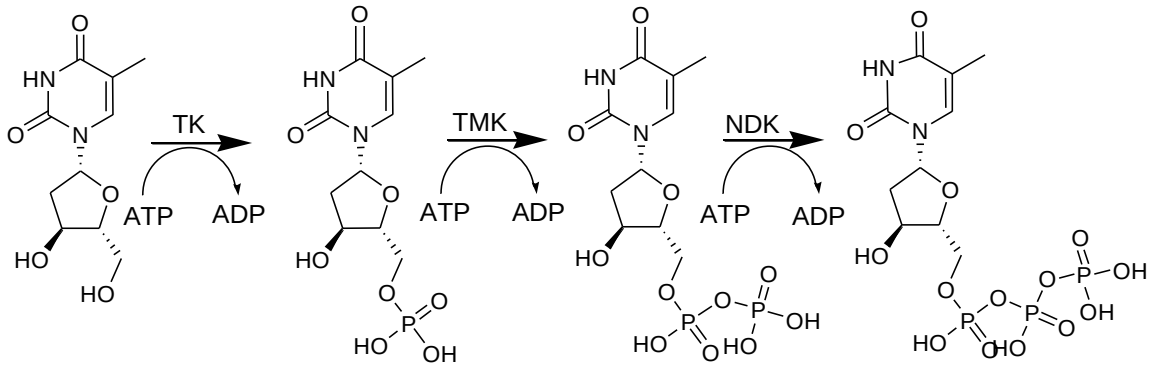
2.3.2 GENERATION OF TDP-L-DIGITOXOSE

The first three steps of the biosynthetic pathway of 6-deoxyhexoses have been characterized by multiple research groups and therefore there was no pressing need to study the first three enzymes of the TDP-L-digitoxose biosynthetic pathway (KijD5, KijD4 and KijB1).[283, 284] For simplicity, we utilized previously characterized enzymes from the rhamnose and mycarose pathways in this work. The gene expression and solubility of these protein products are known, and the purification methodology had previously been developed in our research group. The focus here is on the latter steps of the biosynthetic pathway (**8-9**), as the earlier steps (**4-7**) have been established during studies of previously elucidated pathways. [285, 286, 287]

Native *E. coli* TDP biosynthetic enzymes thymidine kinase (TK), thymidylate kinase (TMK), and nucleotide diphosphate kinase (NDK) were expressed as His_6 tagged

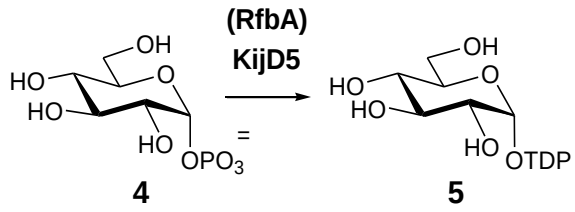
proteins and used to generate thymidine diphosphate (TDP) from thymidine and ATP *in situ* as shown below in Figure 2.6.

2.6 Biosynthesis of TTP from ATP using TK, TMK and NDK



The RfbA protein, a well-characterized protein from the native *E. coli* rhamnose pathway was used instead of the KijD5 protein to couple α-D-glucose-1-phosphate (**4**) to the TDP moiety *in situ* to generate TDP-α-D-glucose (**5**) as shown in Scheme 2.7. The sugar (**5**) was purified by size exclusion chromatography to remove unreacted starting material (TTP and α-D-glucose-1-phosphate) and other co-substrates used in the reaction (pyruvate, ADP, phosphoenolpyruvate). The purity of the product was assessed by analytical HPLC. Under the gradient conditions used, the product eluted at 34 min. Impurities such as thymidine monophosphate (TMP) and thymidine diphosphate (TDP), had HPLC retention times of 25 min and 43 min, respectively.

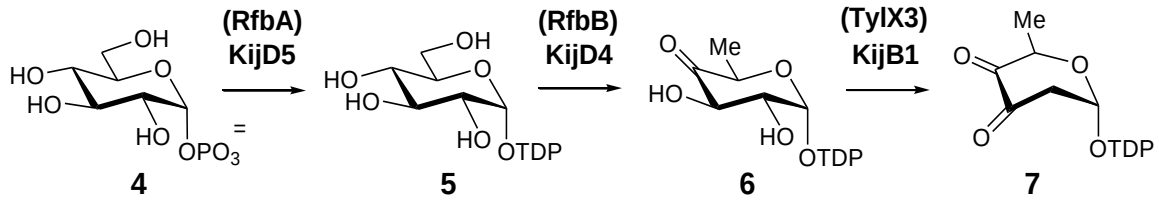
2.7 Biosynthesis of TDP-α-D-glucose from α-D-glucose-1-phosphate (**4**) and TTP



4 **5**

At this stage of purification, it is necessary to remove thymidine monophosphate (TMP) and thymidine diphosphate (TDP) because previous research had shown that TDP is a competitive inhibitor of enzymes involved in the subsequent biosynthetic assays.[288] On a more practical note, TDP is used as the chromophore for our spectrophotometric quantitation, thus excess TDP skews the concentration calculations.

2.8 Early Reactions in the TDP-L-digitoxose pathway

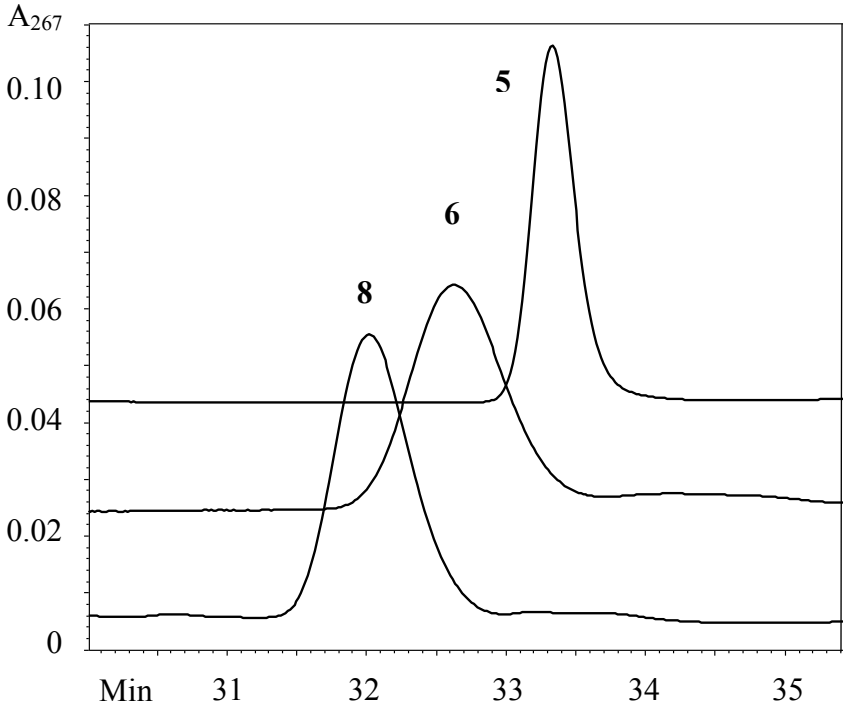


4 **5** **6** **7**

With starting material (**5**) in hand, we chose to employ the RfbB protein, a well-characterized 4,6-dehydratase from the *E. coli* rhamnose pathway for the next conversion step in lieu of the native KijD4 protein. The RfbB protein has been used previously with great success to generate TDP-4-keto-6-deoxy-α-D-glucose (**6**) from TDP-α-D-glucose (**5**). This biosynthetic step has been well characterized by multiple research groups and is the last common intermediate in many sugar biosynthetic pathways.[285] Accordingly,

substrate (**5**) was incubated with enzyme for 3 h at 30 ºC, and the conversion to (**6**) was > 95% complete as assessed by analytical HPLC. A new product peak appeared during the HPLC elution at 33 min, with concomitant decrease of the substrate peak at 34 min.

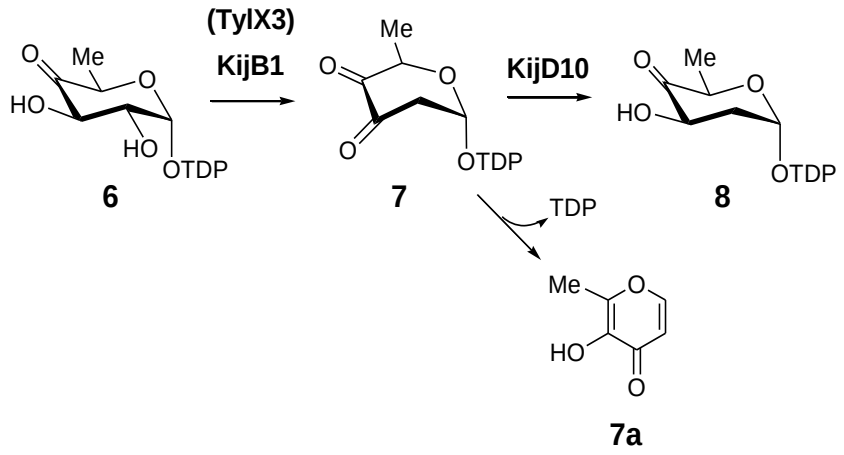
2.7 Overlaid HPLC Traces showing TDP-α-D-glucose (**5**), TDP-4-keto-6-deoxy-α-D-glucose (**6**), and TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**)



As shown in Scheme 2.9, the KijB1 analog, TylX3, a 2,3-dehydratase from the mycarose biosynthetic pathway of *S. fradiae* was incubated with TDP-4-keto-6-deoxy-α-D-glucose (**6**) to generate TDP-3,4-diketo-2,6-dideoxy-α-D-glucose (**7**). The product of this reaction is unstable and has been shown to degrade rapidly to maltol (**7a**) at room temperature, which is a mild condition compared to the 30 ºC environment encountered in antibiotic producing strains of *Streptomyces*. Thus, a coupled assay to generate a more

stable product was carried out with the addition of KijD10 (3-ketoreductase) and NADPH to generate the stable intermediate TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**). [286]

2.9 TDP-4-keto-6-deoxy-α-D-glucose (**6**), conversion to TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**)

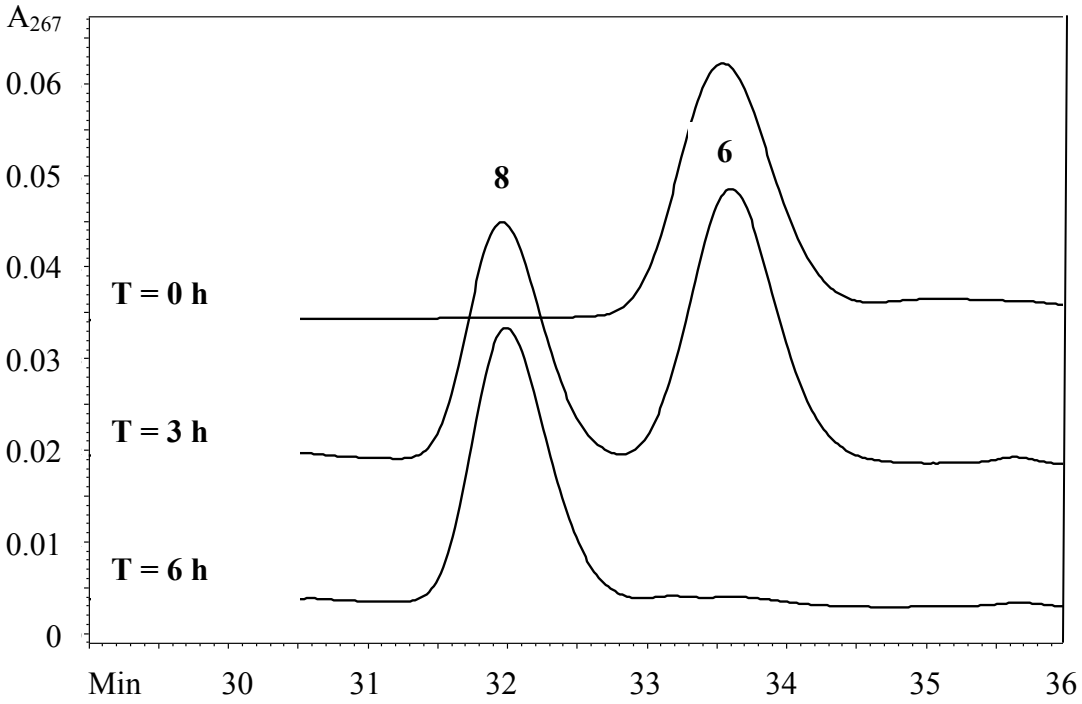


If, as we predict, KijD10 is responsible for the 3-ketoreduction step, we expect to see a $NADP^+$ peak appears in the HPLC trace and increases over time with concomitant increase of a new product peak. We also expect to see a product peak with a retention time of 32 min, as observed for the authentic standard, an intermediate generated in the spinosyn biosynthetic pathway. [287]

To drive the reaction forward, 4 equivalents of KijD10 were added to the reaction for every equivalent of TylX3 protein. The reaction was incubated at room temperature to prevent excess degradation of the unstable TylX3 intermediate (**6**) for 3-6 h or until the reaction reached >95% completion. This produced a new, more stable product (**8**), as demonstrated by a new HPLC peak with a retention time of 32 min as expected, while the

substrate peak (retention time 33 min) decreased in size.(Figure 2.8) These results are in agreement with reported HPLC retention time for TDP-4-keto-2,6-dideoxy-D-glucose.[287]

2.8 Conversion of TDP-4-keto-6-deoxy-α-D-glucose (**6**), to TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**)



Incubation mixture containing 10 mM TDP-4-keto-6-deoxy-α-D-glucose (**6**), 160 nM TylX3, 360 nM KijD10, 1.2 mM NADPH in 50 mM Tris•HCl, pH 7.5, 10% glycerol after 0 h, 3 h, and 6 h, respectively showing time dependent formation of **8** with concomitant decrease of **6**.

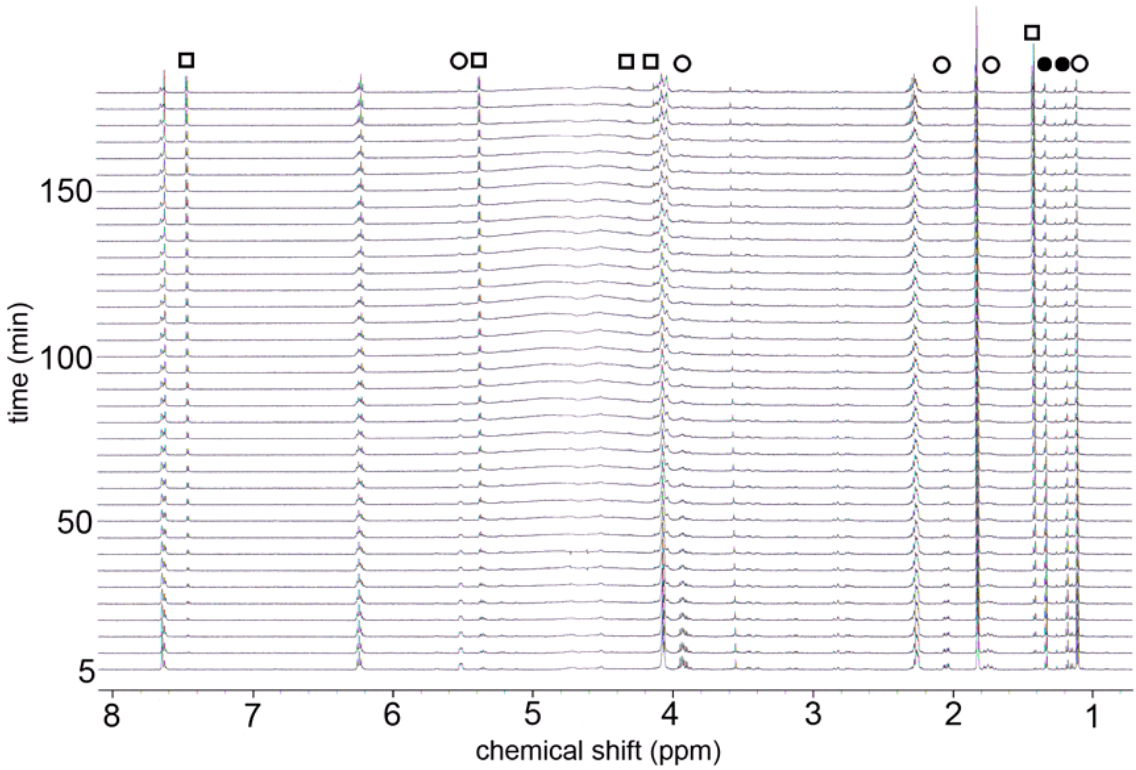
The KijD10 product was characterized by ^{1}H NMR and ^{31}P NMR, and found to be identical to the previously characterized TDP-4-keto-2,6-dideoxy-α-D-glucose generated in the study of the spinosyn pathway from *Saccharopolyspora spinosa*.[287] It

should be noted that as the KijD10 enzyme is subjected to multiple freeze-thaw cycles, the efficiency of conversion drops from >95% in 3 h to 85% over 12 h.

The product (**8**) was purified and subsequently incubated with KijD11, the putative 5-epimerase at room temperature. A new HPLC peak (**9**) with a retention time of 35 min appeared and increased over time, while the substrate peak showed concurrent decrease. As the reaction progressed, another peak appeared with the retention time of approximately 2 min as the 35 min product peak (**9**) began to decrease. This reaction was repeated for *in situ* ^{1}H-NMR analysis, which showed the production of two compounds; a predicted product (**9**) and a compound appearing to result from degradation of **9** (**9b**, Figure 2.10). It is likely that the epimerized product, TDP-4-keto-2,6-dideoxy-L-allose (**9**), degrades to a pyran ring and TDP. (See Figure 2.10)

The structure of the degradation product cannot be fully assigned from this NMR spectrum due to the complexity of the reaction mixture. Those signals corresponding to the KijD11 product, TDP-4-keto-2,6-dideoxy-L-allose (**9**), reached maximum intensity at ~20 min and then diminished.[289] A third set of signals corresponding to the degradation product appeared after a short lag. The ^{1}H-NMR and COSY spectra taken at the end of reaction allowed the verification of (2S, 3S)-2-methyl-3-hydroxy-4-keto-2,3-dihydropyran as the degradation product.[290] As demonstrated by HPLC analysis and NMR *in situ* assay, degradation of the intermediate (**9**) begins shortly after the reaction commences, therefore to generate the desired product a coupled assay was performed. Due to the apparent instability of the intermediate, the KijD11 and KijC2 proteins were incubated with the TDP-4-keto-2,6-dideoxy-α-D-glucose substrate. A new peak with a retention time of 34 min appeared in the HPLC trace (**10**). (See figure 2.11)

2.9 ^{1}H NMR Time Course Analysis of the KijD11 Reaction

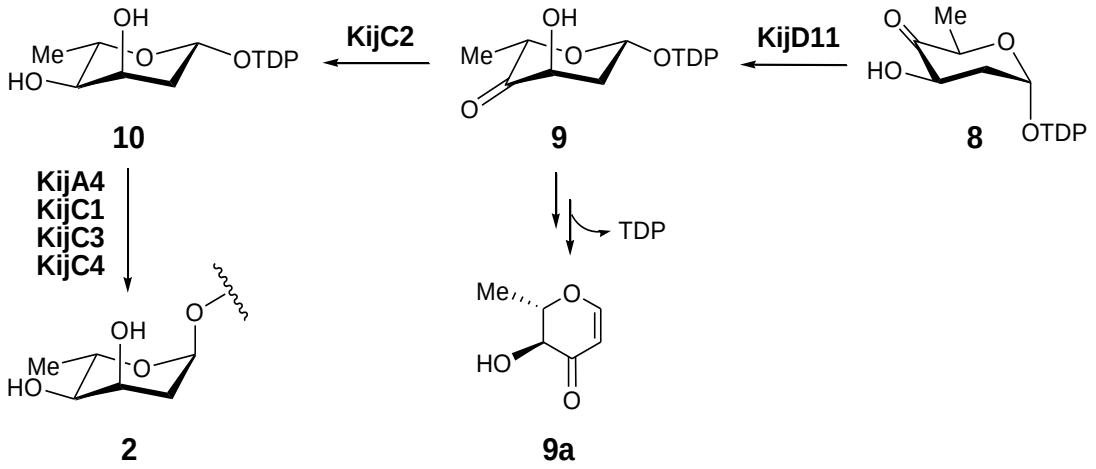


^{1}H NMR stack plot of the KijD11 reaction [10 mM (**5**) and 10 μM KijD11 in 50 mM KH_2PO_4 buffer (pH 7.5)] monitored over 180 min. The signals corresponding to each compound are labeled: (O) from **8**,(●) from **9**, and (□) from **9a** (see Experimental Procedures for details).

TDP-L-digitoxose is a highly unstable compound and degrades during its biosynthesis. Attempts at purification using methods normally applied to these sugars, such as size exclusion chromatography using Bio-Gel P2 resin failed. When the TDP-L-digitoxose was applied to the column, absorption profile of the elution fractions indicated

that the expected product had degraded to an uncharacterized product, most likely a pyran ring.

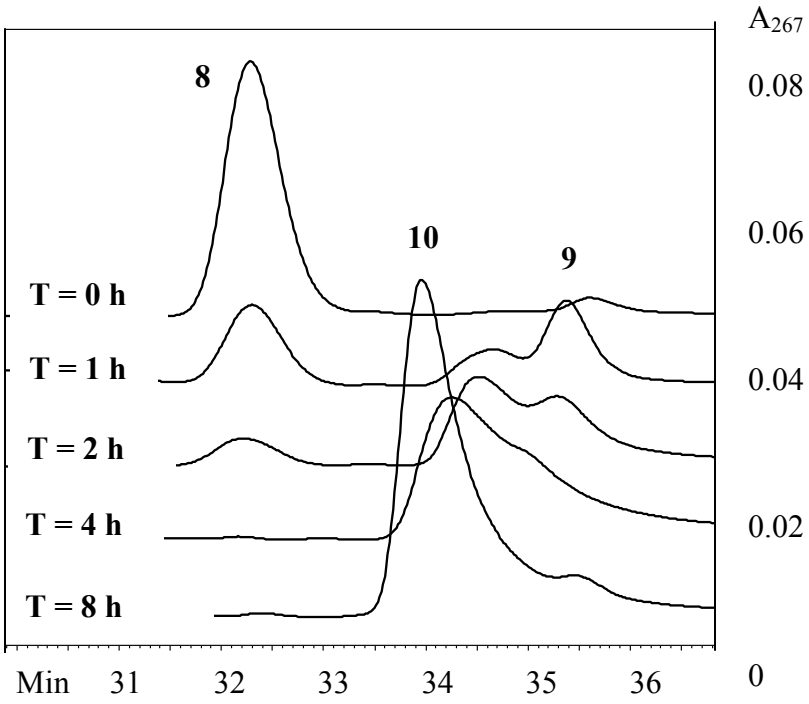
2.10 Proposed Conversion of TDP-4-keto-2,6-dideoxy-L-allose (**9**) to a Pyran Ring Product (**9a**) and TDP



Our only option to purify the compound was therefore to employ FPLC using ion exchange chromatography, however the high salt concentrations also proved to be problematic for the stability of the product, as concentration in the presence of bicarbonate appeared to cause additional degradation of the product. Thus, we concentrated the sugar to a small volume and applied it to a very short P2 column in order to remove as much of the bicarbonate buffer as possible.

The fractions found to contain the sugar were additionally desalted using isocratic elution using an HPLC equipped with a C-18 column, with water as the eluting solvent. This step also served to remove remaining glycerol and dithiothreitol in the sample in preparation for NMR analysis.

2.11 Analytical HPLC Traces of TDP-L-digitoxose (**10**) Conversion from TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**)

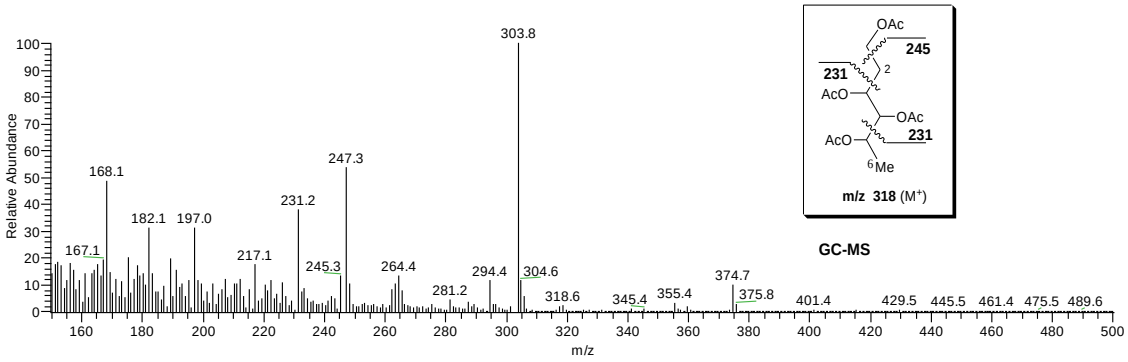


Incubation mixture containing 1 mM TDP-4-keto-2,6-dideoxy-α-D-glucose (**8**), 100 nM KijD11, 10 μM KijC2, 1.4 mM NADH and 10 mM $MgCl_2$ in 50 mM Tris•HCl, pH 7.5, 10% glycerol after 1 h, 2 h, 4 h, and 8 h, respectively showing time dependent formation of **10**

The fractions containing sugar were lyophilized again and stored at -80 ºC. Unfortunately, the sugar was not stable enough to characterize using NMR, and our attempts at high resolution mass spectrometry failed due to the instability of the TDP-sugar. We were able to obtain low resolution mass spectrometry results using ESI-MS and CI (negative mode). The calculated mass for the expected product, TDP-L-

digitoxose, is 532 and the found mass was in CI was 532, and the found mass was 265 in ESI-MS, which corresponds to the -2 charge state. Because the product is too unstable to be characterized by NMR, we decided to generate the TDP-sugar, and obtain enough material for chemical acetylation in order to characterize the product.

2.12 GC-MS of TDP-L-Digitoxose Acetylation Products



A large-scale synthesis was carried out from 5 mg of compound (**8**) following the method presented earlier in this chapter. After the synthesis, the product (**10**) was purified using FPLC only, and fractions were combined and subjected to acid hydrolysis following the procedure reported by Blakeney et al and previously used in our research group. [281, 282] After the TDP-moiety was removed from the TDP-L-digitoxose (**10**), the resulting compound, (**11**) was dried and reduced with sodium borohydride, followed by peracetylation using 1-methyl-imidazole and acetic anhydride. The final product, tetraacetyldigitoxose (**12**) was analyzed and allowed us to confirm that these gene products together produce TDP-L-digitoxose.

2.4 CONCLUSIONS

Herein, we report the results of our investigations of the function of the genes of the TDP-L-digitoxose biosynthetic pathway. We have confirmed the assignments of the genes as follows: KijB1 encodes a 2,3-dehydratase, KijD10 encodes the 3-ketoreductase, KijD11 encodes a 5-epimerase, and KijC2 encodes the 4-ketoreductase. (See schemes 2.9 and 2.10) Previously, *in vitro* assays for testing the activity of the glycosyltransferases have been hampered by the unavailability of TDP-L-digitoxose. Our success in reconstructing the deoxysugar pathway allows further investigation into the role and specificity of each of the glycosyltransferases that build the digitoxose tetrasaccharide chain of kijanimicin. We have now identified the genes necessary to generate TDP-L-digitoxose *in vivo* and *in vitro*, thus this portion of the research allows us to investigate the activity of the kijanimicin glycosyltransferases that were identified through gene cluster sequencing.

There are no glycosyltransferases known to date that can accept kijanolide as a substrate, which provides a new aglycone scaffold for combinatorial biosynthesis, thus with the elucidation of this pathway, we have not only confirmed the identity of the kijanimicin biosynthetic gene cluster, but we are now also able to investigate the activity of these novel glycosyltransferases in the kijanimicin gene cluster. To harness the full potential of the combinatorial approach for therapeutic purposes, a glycosyltransferase "gene toolbox" must be generated, which should include glycosyltransferases that recognize different, structurally diverse antibiotic aglycones.[291] As reported in a later chapter, our research expands this toolbox, by providing four novel glycosyltransferases for use in combinatorial biosynthesis that may demonstrate interesting and unique substrate flexibility or glycosylation activity.

,

3 Chapter 3: Prokaryotic Glycosylation Reactions: *Reconstruction of the Glycosylation Reactions of Kijanimicin Biosynthesis*

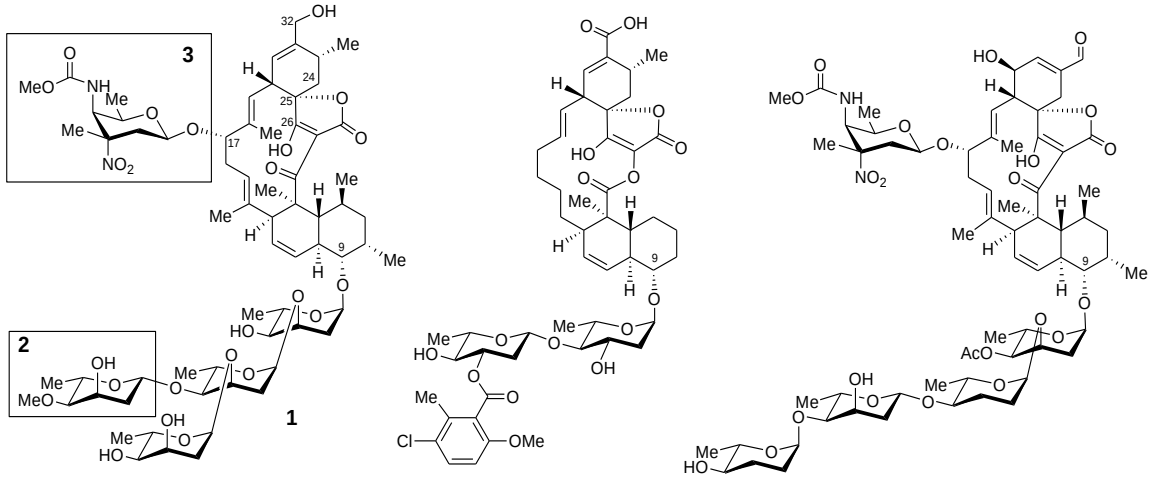
3.1 INTRODUCTION

It has been argued that glycosyltransfer reactions are, on quantitative terms, the most important biological transformations in Nature, since they account for the biosynthesis and hydrolysis of the bulk of biomass.[3] As a logical extension, one may also argue that the most important enzymes are those that are involved in glycosylation reactions, including unusual sugar biosynthesis, glycosyltransferases which catalyze the transfer of sugar moieties from activated sugar donor molecules to specific acceptor molecules to form glycosidic bonds, and glycosidases, which catalyze the removal of sugar moieties from glycosylated acceptor molecules. Notably, the biosynthesis of disaccharides, oligosaccharides and polysaccharides involves the action of literally hundreds of different glycosyltransferases.[4] The importance of glycosylation to natural product biosynthesis is no exception to this argument, and in fact may serve to demonstrate the importance of glycosylation as the bioactivity of many natural products including clinically important antibiotics and anticancer therapeutics depends on their sugar moieties. [292]

The investigation of kijanimicin (**1**) glycosylation is particularly interesting, due to the presence of a branched-chain tetrasaccharide moiety consisting of three units of 2,6-dideoxy-α-L-hexopyranose (digitoxose) and one unit of 2,6-dideoxy-4-*O*-methyl-*β*-L-hexopyranose (digitoxose) (**2**), as well as the novel amino sugar kijanose (**3**) as shown in Figure 3.1.[246] Ubiquitous in nature, 2,6-dideoxysugars such as digitoxose are especially abundant in bacterial secondary metabolites, including macrolide, polyene,

enediyine, pregnane and aminoglycoside antibiotics, while kijanose is quite rare.[241] This nitro sugar had been found in only a few spirotetronate antibiotics, such as kijanimicin (**1**), lobophorin B, tetrocarcins, and arisostatin A. [68, 249, 293-295] The unusual nitrosugar D-kijanose (**3**), whose biosynthesis likely involves new chemistry, is still under investigation by multiple research groups.[68] A survey of known spirotetronate natural product structures shows diverse glycosylation patterns, particularly in the oligosaccharide chain attached to the C-9 position. The structure of the oligosaccharide chain at the C-9 position of tetrocarcin has been shown to be an important modulator of its Bcl-2 inhibition activity, suggesting that glycodiversification at this position may lead to spirotetronate derivatives with high therapeutic value. [296]

3.1 Tetronolide Natural Products: Highlighting Differential Glycosylation at C-9



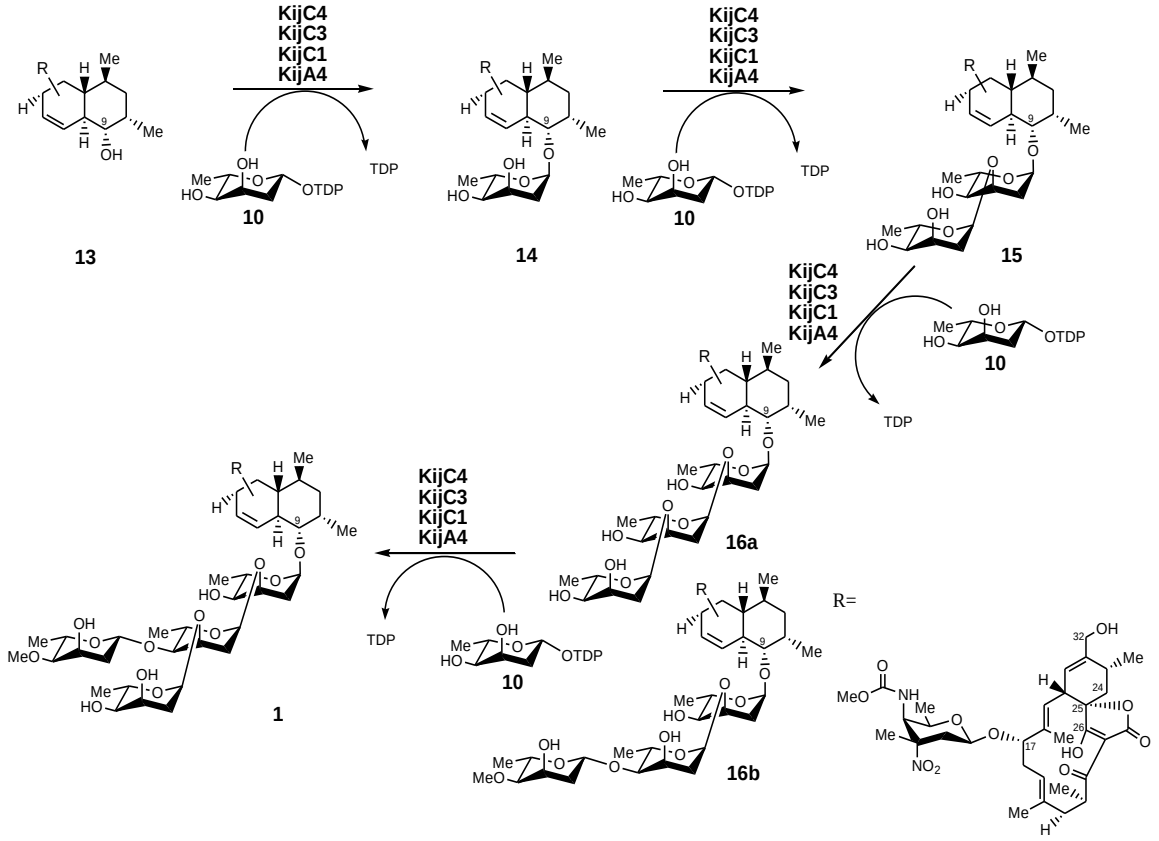
Landomycin A is a hexasaccharide-containing compound which is produced by *Streptomyces cyanogenus* S13622 and exhibits strong antitumor activities, in particular

against prostate cancer lines.[71] These unusual activities can be attributed to the presence of the long oligosaccharide chain, since landomycins with shorter sugar side chains as landomycin E produced by *S. globisporus* possesses much weaker antitumor activity.[71] Thus, we are interested in identifying the glycosyltransferases responsible for the generation of the C-9 glycosylation long chain oligosaccharide. The biosynthesis of TDP-L-digitoxose was elucidated in chapter 2, thus enabling experimental analysis of kijanimicin digitoxosyltransferase activity. The presence of TDP-L-digitoxose in biologically active natural products with varied functions makes digitoxose an attractive building block to be used in combinatorial biosynthesis to glycosylated secondary metabolites, however combinatorial biosynthesis requires not only the NDP-sugar donors, but glycosyltransferases that can accept these NDP-sugars as substrates. To discover the glycosyltransferases that are responsible for the addition of TDP-L-digitoxose to the tetronolide aglycone, the biosynthetic gene cluster of kijanimicin was examined *in silico*.

Using NCBI Blast, open reading frames in the kijanimicin biosynthetic cluster were examined and 5 glycosyltransferase genes were identified. Sequence analysis demonstrated that KijD9 has 38-40% pairwise identity to other glycosyltransferases in the cluster. The remaining four glycosyltransferases KijC1, KijC3, KijC4 and KijA4 exhibit 50-58% pairwise identity to each other as well as to the chlorothricin glycosyltransferases ChlC6 and ChlC7. [68, 293] Chlorothricin is another spirotetronate natural product decorated with 2 digitoxose moieties. Thus the strong similarity of the KijGTs to the ChlGTs suggests that they share similar aglycone and sugar substrates.[293] Taken together, these data strongly suggest that KijC1, KijC3, KijC4 and KijA4 are digitoxosyltransferases responsible for constructing the C-9 oligosaccharide of kijanimicin.[68] From sequence alignment, it is not possible to predict the order of the

glycosylation events, nor which protein is responsible for a specific digitoxosyl attachment.

3.1 Proposed Glycosylation Pathway in Kijanimicin Biosynthesis



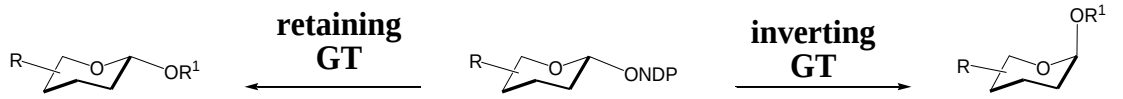
Most GTs involved in natural product biosynthesis catalyze the transfer of one sugar to one glycosidic bond.[297] There exist only a few examples of bifunctional GT enzymes responsible for the catalysis of two glycosidic bonds in the literature (e.g., LanGT1, LanGT4 and AveBI). [298, 299] Thus, we proposed a scheme involving four

glycosyltransferases in the sequential transfer of four digitoxosyl moieties in the biosynthesis of kijanimicin as shown in Scheme 3.1.

The order of transfer is unknown and the latter steps, especially the order of attachment of the third and fourth sugars, can only be determined experimentally. If the β(1-4) linkage is formed before the terminal α(1-3) linkage, the resultant 3-sugar compound will be **16b**, if the obverse is true, the resulting 3-sugar intermediate compound will be **16a**. In any event, the final addition, be it the α(1-3) or β(1-4) linkage, the end product should be kijanimicin A (**1**).

Returning to focus on the *in silico* characterization of the kijanimicin glycosyltransferases, these proteins share sequence homology and align to glycosyltransferases that are members of the GT-B superfamily. Proteins of the GT-B superfamily are bi-lobal enzymes that have two β/α/β Rossmann-like domains with the active site lying in the cleft between the domains.[79] Each lobe of the enzyme is associated with either the NDP-sugar or the aglycone substrate.[79] It should be noted that enzymes in the same superfamily or having the same folds so not necessarily have similar mechanisms or the same stereochemical outcome of the product.[80] Indeed, non-glycosyltransferase enzymes are also known to adopt the GT-B fold, an example of which is UDP-GlcNAc 2-epimerase. [81] Notably, the kijanimicin glycosyltransferases, which belong to the GT-B family exhibit interesting stereochemical variance, since they are responsible for creating both α and β linkages from a TDP-sugar substrate that originally has a β configuration. Thus, three glycosyltransferases in this pathway must catalyze an inverting (S_N2) reaction and one must catalyze a retaining or (S_N1) glycosyltransfer reaction.

3.2 Inverting and Retaining Glycosyltransfer Reactions for an L-sugar

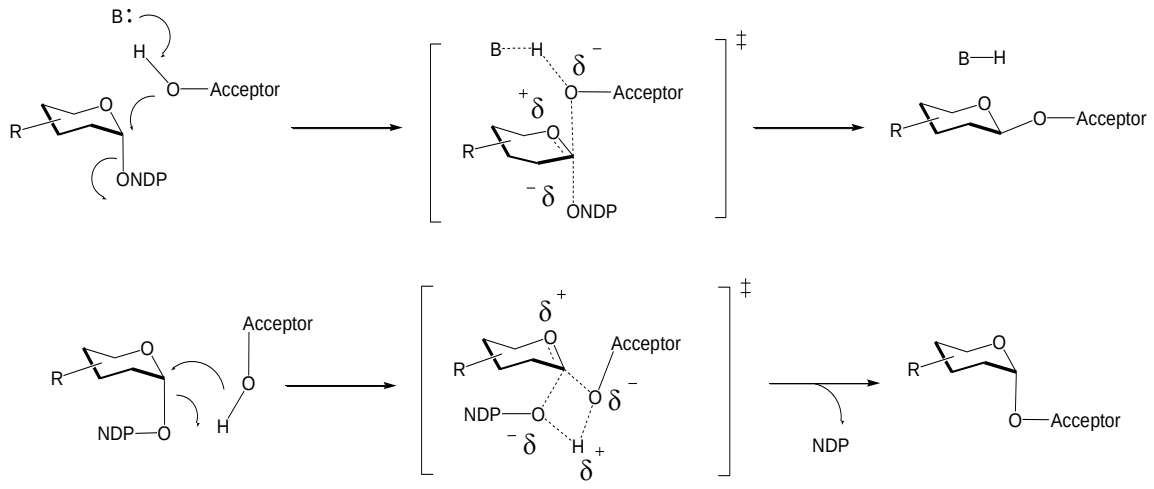


β-L-glycoside NDP-β-L-sugar α-L-glycoside

The members of the GT-B superfamily seem to display a higher degree of diversity in the catalytic mechanisms as compared to what has been observed for the inverting GT-A enzymes. [79] The greater mechanistic diversity is perhaps a reflection of the greater diversity of reactions catalyzed by this superfamily of enzymes. This may also be a result of the greater physical separation of the two domains in the enzymes. Therefore, the mechanisms of these enzymes cannot be definitively predicted, although a general mechanism can be described.

Briefly, the S_N2 inverting mechanism involves the direct attack at the anomeric center by the acceptor, displacing the NDP activating group. In this mechanism, the active site nucleophile, is proposed to stabilize and position the aglycon in the active site for the direct attack, while the NDP group is displaced with the assistance of a metal ion.[79] As the NDP group is displaced a basic residue within the enzyme active site abstracts a proton from the aglycone hydroxyl. [79] The resulting hydroxyl anion can then attack the opposing face of the sugar in a concerted manner, finally transferring the sugar to the aglycone with inversion of the stereochemistry of the anomeric center. Thus, the nucleophilic acceptor attacks the anomeric carbon from the opposing side of the sugar ring as the NDP leaving group in a synchronous, concerted manner with highly dissociative oxocarbenium-like character. [82]

3.3 Mechanisms of Inverting and Retaining Glycosyltransferases



Meanwhile, the S_N1 retaining mechanism is proposed to pass through a short-lived oxocarbenium transition state, which is facilitated by the lone pair electrons from the endocyclic oxygen. As the NDP group displacement occurs without the stabilization of the active site nucleophile, a transient oxocarbenium ion forms, and a proton is abstracted from the aglycone acceptor hydroxyl. The formation of the oxocarbenium ion enables the direct attack of the anomeric carbon by the aglycon hydroxyl ion, retaining the stereochemistry of the original NDP group. [83] These mechanisms are shown in Figure 3.3. The specific mechanisms of the kijanimicin glycosyltransferases cannot be established until their functions are elucidated through experimental analysis, however it is unlikely that they are vastly divergent from the mechanisms presented above.

The focus of this research is to elucidate the order of attachment of digitoxosyl moieties to the kijanolide aglycone to generate the tetrasaccharide chain. The branched oligosaccharide chain has been previously synthesized by Thiem and Kopper in 1990

using an *N*-iodosuccinimide procedure employing regioselectively blocked L-digitoxosides and L-digitoxals. [300] The tetrasaccharide chain was later characterized extensively by ^{1}H-NMR. This work will attempt to isolate and characterize any intermediate products that may be generated during these studies. No previous biosynthetic studies have been carried out to investigate the formation of the branched tetrasaccharide, despite evidence that modification of the C-9 glycan may have a significant effect on the bioactivity of the natural product.

Our research plan includes the cloning and expression of the five glycosyltransferase genes in *S. lividans*. The kijanimicin aglycone will be isolated and used in assays both *in vivo* and *in vitro* to build the tetrasaccharide chain one sugar "link" at a time. Each of the intermediate products will be isolated and characterized, then used as substrates for subsequent reactions until kijanimicin A is constructed. Once the digitoxosyltransferases have been identified and their activity confirmed, potential glycodiversification experiments can be carried out to alter the C-9 glycosylation pattern of kijanimicin. This alteration could yield new natural products with different bioactivities suitable for potential clinical applications.

3.2 EXPERIMENTAL DETAILS

3.2.1 MATERIALS

3.2.1.1 Bacterial Strains

Escherichia coli (*E. coli*) strain DH5α, *Streptomyces lividans* TK-24 strain and *Actinomadura kijaniata* SCC1256 (ATCC 31588) were purchased from Bethesda

Research Laboratories (Gaithersburg, MD) and the American Type Culture Collection (Manassas, VA), respectively. The overexpression hosts *E. coli* BL21, BL21(DE3), BL21(DE3)PLysS and BL21 RosettaII(DE3) were purchased from Novagen (Madison, WI). *E. coli* XL-1 Blue MRF' used for cosmid library preparation was obtained from Stratagene (La Jolla, CA).

3.2.1.2 Biochemicals

Enzymes including and molecular weight standard used for the molecular cloning experiments were products of Invitrogen (Carlsbad, CA) or New England Biolabs (Beverly, MA). Restriction digestion enzymes, calf intestinal alkaline phosphatase (CIAP), 100X BSA, T4 ligase, and their respective buffers were products of New England Biolabs (Beverly, MA). Ni-NTA agarose resin was obtained from Qiagen (Valencia, CA). Growth media components were acquired from Becton Dickinson (Sparks, MD). Rabbit muscle pyruvate kinase was purchased from Sigma as a 400-800 unit/mg ammonium sulfate precipitate. This ammonium sulfate precipitate was dissolved in water to a concentration of 2500 units/mL, dialyzed against 50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0, to remove ammonium sulfate, and stored at -80 °C. Antibiotics and chemicals such as isopropyl-β-D-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were products of Sigma-Aldrich Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). Agarose for DNA electrophoresis was obtained from Fisher Scientific (Pittsburgh, PA). Bio-Gel P2 resin and all reagents for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were purchased from Bio-Rad (Hercules, CA), with the exception of the prestained protein molecular weight marker, which was ordered from New England Biolabs. Antibodies used for Western blotting procedures, monoclonal

anti-polyhistidine clone His-I H1029 and anti-mouse IgG (alkaline phosphatase conjugate) A2429 were purchased from Sigma. For Western blot detection, nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) were both obtained from Promega.

3.2.1.3 Plasmids, Vectors and DNA Manipulation

Vector pET28b(+) was purchased from Novagen (Madison, WI). Gigapack III Gold packaging extract, Plasmids SuperCos 1 and pBluescript II SK(+), used for cosmid library construction and cosmid fragment subcloning, respectively, were purchased from Stratagene (La Jolla, CA). DNA gel extraction and spin miniprep kits were obtained from Qiagen (Valencia, CA). DNA polymerase *pfu* was purchased from Stratagene. Oligonucleotide primers for cloning were prepared by Integrated DNA Technologies (Coralville, IA). Plasmid vectors pSET152, pWHM3, pWHM468 and pWHM467 were a kindly provided by Dr. C. R. Hutchinson.[301-303] The pLitmus28 vector was obtained from NEB. pET28 a and b+ vectors were purchased from Novagen.

3.2.1.4 Instrumentation

PCR reactions were carried out using an Eppendorf Mastercycler gradient thermal cycler from Brinkman-Eppendorf (Westbury, NY). For agarose gel electrophoresis, a mini-sub-cell GT from BioRad (Richmond, CA) was used, powered by a Fisher FB-300 electrophoresis power source. After electrophoresis, DNA bands were visualized using a Fisher FBTIV-08 UV transilluminator. Cell disruption was performed with a Fisher 550 Sonic dismembrator, equipped with a standard sonicator horn. For SDS-page gel electrophoresis, a mini PROTEAN II vertical system from BioRad (Richmond, CA) was

used, equipped with an FB-300 or EC-1000-90 electrophoresis power source from Fisher, or E-C apparatus corporation, respectively. Mini-trans blot assembly apparatus used for Western blot, GelAir drying system for acrylamide gel preservation, and the necessary accessories and reagents were all products of BioRad. The nitrocellulose Hybond-C or Optitran BA-5 85 reinforced nitrocellulose membrane used for Western blot was from Amersham Bioscences or Whatman (Dussel, Germany), respectively.

To obtain pH values, a Corning 240 pH meter equipped with an Accumet electrode (Fisher) was used. Large scale centrifugation procedures were performed using an Avanti J-25 unit or a J-E instrument from Beckman (Arlington Heights, IL). Microscale centrifugation procedures were carried out using an Eppendorf 5415C from Brinkmann Instruments, (Westbury, NY). Ultraviolet-visible spectra were recorded using Beckman DU-650 spectrophotometer. Amicon YM-10 filtration products were purchased from Millipore (Billerica, MA). HPLC separations used a Beckman 366 from Beckman Instruments (Fullerton, CA) equipped with a CarboPac PA1 HPLC column from Dionex (Sunnyvale, CA) or a C-18 from Varian (Lake Forest, CA). Polygram Sil G/UV_{254} plates, 0.25 mm; were purchased from Macherey-Nagel Inc. (Easton, PA). DNA and compounds were dried using a Savant Speedvac SC100, equipped with a Savant refrigerated condensation trap RT100 and Gel Pump GP100. DNA concentrations were measured using a NanoDrop ND-1000 UV-Vis instrument from Thermo Fisher Scientific (formerly NanoDrop Technologies LLC).

NMR spectra were acquired on either a Varian Unity 300 or a 500 MHz spectrometer. Sequencing of subclones was performed using a capillary-based AB 3700 DNA analyzer by the Core Facilities of the Institute of Cellular and Molecular Biology at the University of Texas at Austin. Mass spectra (low-resolution ESI) were obtained on a Finnegan LCQ ion trap mass spectrometer. High-resolution ESI spectra were obtained

using an IonSpec 9.4T FT-ICR mass spectrometer. Spectra were acquired by the Mass Spectrometry Core Facility in the Department of Chemistry and Biochemistry of the University at Texas at Austin.

3.2.2 General Procedures

Protein concentrations were determined according to Bradford [272] using bovine serum albumin as the standard. The relative molecular mass and purity of enzyme samples were determined using SDS-polyacrylamide gel electrophoresis as described by Laemmli.[273] NMR chemical shifts (δ in ppm) are given relative to that of deuterated water (D_2O, δ 4.65 for ^1H NMR), with coupling constants reported in Hertz (Hz).

3.2.3 Experimental Procedures

3.2.3.1 Bacterial Strains and Growth Conditions

Actinomadura kijaniata SCC1256 (ATCC 31588) was maintained at 30 ºC on ISP medium 2. For kijanimicin production, spore suspensions of wild-type *A. kijaniata* were grown in seed media (0.3% beef extract, 0.5% tryptose, 0.1% dextrose, 2.4% potato starch, 0.2% $CaCO_3$, pH 7.2) and vegetative media (0.5% yeast extract, 1% dextrose, 2% soluble starch, 0.5% NZ-amine A, 0.4% $CaCO_3$, 0.024% $CoCl_2 \cdot 6H_2O$).[243] In both seed and vegetative media, *A. kijaniata* was grown on a rotary shaker at 30°C, 250 RPM.

S. lividans TK-24 strain was maintained on R1R2 media (10.3% sucrose, 1% glucose, 1% $MgCl_2$, 0.5% yeast extract, 2.2% agar, 0.15% L-proline, 0.1% L-asparagine, 0.01% casamino acids, 0.005% KH_2PO_4, 0.3% $CaCl_2$, 10% 0.25 M TES buffer, pH 7.2, 0.2% trace elements).[276] Trace elements solution is composed of: 0.004% $ZnCl_2$, 0.2% $FeCl_3 \cdot 6H_2O$, 0.001% $CuCl_2 \cdot 2H_2O$, 0.001% $MnCl_2 \cdot 4H_2O$, 0.001% $Na_2B_4O_7 \cdot 10H_2O$, and

0.001% $(NH_4)Mo_7O_{24}$•$10H_2O$. Liquid cultures of *S. lividans* TK-24 employed YEME media (34% sucrose, 1% glucose, 0.5% bacto-peptone, 0.3% yeast extract, 0.3% malt extract).[276]

Escherichia coli DH5α was used for routine cloning experiments. *E. coli* BL21, BL21(DE3), BL21(DE3)pLysS, and BL21 RosettaII(DE3) were cultured in LB-media supplemented with 100 μg/mL ampicillin or 50 μg/mL kanamycin with the addition of 35 μg/mL chloramphenicol in the case of BL21 RosettaII(DE3) and BL21(DE3)pLysS . To liquid cultures, a volume of sterile 80% glycerol equal to 20% of the original culture volume was added and gently mixed to homogeneity. The mixture was aliquotted into sterile Eppendorf tubes, flash frozen in liquid N_2 and stored at -80 ºC until use. Glycerol stocks were used in a 1:20 dilution to inoculate fresh media for plasmid preparations. Competent cells were prepared as described in chapter 2.

3.2.3.2 DNA Manipulation

Template DNA used in the following experiments in this chapter was generated by Dr. Hua Zhang, following the procedures described in Chapter 2. Routine genetic manipulations in *E. coli* were performed as described by Sambrook et al.[274] Standard genetic manipulations in *S. lividans* were performed as described by Kiser et al.[276] Cosmids used in these experiments were: pHZ-8 and pHZ4-6. The ermE* promoter was retrieved as a 279-bp *Kpn*I-*Bam*HI fragment from pWHM79 and ligated into the same restriction sites of pWHM3, which generated pWHM3-ermEp.[304] Subcloning of the DNA fragment harboring actIIorf4 and actI promoter was previously reported.[305]

3.2.3.3 PCR Reactions

The general protocol for amplification of genes from gram-positive bacteria is as follows. Primers were prepared by dilution with TE buffer (50 mM Tris, 1 mM EDTA, pH 8.0) to a concentration of 100 µM. This stock solution was further diluted to 20 µM with sterile water. Template DNA from the *A. kijaniata* genome was obtained from cosmids constructed by Dr. Hua Zhang. Glycosyltransferase genes *kijc1, kijc3, kijc4,* and *kija4* were amplified from cosmid pHZ4-6, while *kijd9* was amplified from cosmid pHZ-8. Polymerase chain reactions (PCR) consisted of the following: 72 µL of sterile water, 10 µL of 10X *pfu* polymerase buffer, 10 µL of dimethylsulfoxide (DMSO), 10 µL of a 20 mM deoxyribonucleotidyl triphosphosphate (dNTP) solution, 2.5 µL of each primer (5′ and 3′) at 20 µM, 1 µL of template DNA, and 2 µL of *pfu* polymerase (5 units), for a 100 µL total reaction volume in an 0.5 mL thin walled PCR tube.

The PCR reactions were carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial denaturing step of 2 min at 95 ºC. The first 3 program cycles consisted of a 30 s denaturing step at 95 ºC, an annealing step at 5 ºC below the lowest annealing temperature for the overlapping region of each primer pair for 30 s, and an elongation step of 1 min/kb at 72 ºC. The remainder of the program consisted of a 30 s denaturing step at 95 ºC, an annealing step at an annealing step with temperature 5 ºC below the lowest annealing temperature (for the entire primer) of each primer pair for 30 s, and an elongation step of 1 min/kb at 72 ºC. This cycle was repeated 27 times, with program ending in a 10 min extension at 72 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

3.2.3.4 PCR Product Purification

A typical PCR purification consisted of agarose gel electrophoresis employing an 0.8% TAE-agarose gel with 5 μg/mL ethidium bromide. PCR products were diluted to 83% with 6X DNA loading dye (0.25% xylene cyanol, 0.25% bromophenol blue, 15% Ficoll 400) and electrophoresed at 100 V for 30 min. After electrophoresis, DNA bands of the correct base pair length, as estimated from comparison with DNA ladder, were excised from the gel with a clean scalpel. DNA was purified from agarose slices using the Qiagen gel extraction kit, following the manufacturer's protocols with the following modifications. Instead of heating at 55 ºC to melt the agarose, gel slices were heated at 65 ºC until the gel was fully dissolved. DNA elution buffer EB was also heated to 65 ºC prior to DNA elution.

3.2.3.5 DNA Ligations and Plasmid Transformation

PCR products for the kijanimicin biosynthetic protein constructs were digested with the appropriate restriction enzymes, generally *Nde*I and *Hind*III at 37 ºC overnight. A typical reaction mixture contained 30 μL of purified PCR product, 1-2 μL of each enzyme, 0.4 μL of 100X bovine serum albumin (BSA), 4 μL of 10X NEB buffer 2 and sterile water to a final volume of 40 μL. Digested products were cleaned with the Qiagen PCR purification kit, following the manufacturer's protocols, and ligated to the corresponding sites in the pET28 a or b+ plasmid. The pET28 a or b+ plasmid was prepared by digesting with *Nde*I and *Hind*III at 37 ºC overnight followed by treatment for 1-2 h at 37 ºC with calf intestinal alkaline phosphatase (CIP) in NEB buffer #2. A typical reaction mixture contained 30 μL of purified vector, 2 μL of each enzyme, 0.4 μL of 100X BSA, 4 μL of 10X NEB buffer 2 and sterile water to a final volume of 40 μL. After overnight digestion, pET28a or b+ was treated for 1-2 h at 37 ºC with 0.2 μL of CIP in

NEB buffer #2. The digested plasmid was electrophoresed using a 0.8% TAE-agarose gel, followed by extraction using the Qiagen gel extraction kit with modified protocols as previously described. In some cases, plasmids were extracted using the Qiagen gel extraction kit, following the instructions for PCR purification.

A Nanodrop instrument was used to quantify the plasmid and insert DNA solutions. A typical ligation mixture included 90 fm ends of gel purified PCR product, 30 fm ends of pET28a or b+, 2 µL of 5X T4 ligase buffer, 3.5 µL of sterile water, and 0.5 µL of T4 ligase to a final volume of 10 µL. The ligation mixture was incubated at room temperature for 1-2 h, or overnight at 4 ºC. An aliquot of the ligation mixture (5 µL) was introduced to competent *E. coli* DH5α cells (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, and incubated on ice for 2 min. LB broth (500 µL) was added to the transformed cells and incubated at 37 ºC with 200 RPM shaking for 45 min. After incubation, transformation reactions were plated on LB-agar plates supplemented with 50 µg/mL kanamycin. Positive transformants were selected after incubation overnight at 37 ºC.

3.2.3.6 DNA Extraction by Alkaline Lysis

Colonies appearing after overnight incubation were screened for the presence of plasmid containing PCR product using the method of alkaline lysis.[277] Briefly, 5 mL cultures of each positive transformant were grown overnight at 37 ºC in LB media supplemented with 50 µg/mL kanamycin. Cultures were centrifuged at 12,000 *g* for 30 s, the supernatant was removed and the cells were resuspended in 100 µL of cold Resuspension solution (50 mM glucose, 25 mM Tris•HCl (pH 8.0), 10 mM EDTA (pH 8.0), 100 µg/mL RNase). Cells were disrupted with the addition of 200 µL fresh Lysis

solution (200 mM NaOH, 1% SDS), and inverted five times to mix. Lysis reactions were quenched with the addition of 150 µL of Neutralization solution (3 M potassium acetate, 11.5% acetic acid). Reactions were mixed to ensure neutralization and incubated at 4 ºC for 5 min, followed by centrifugation for 5 min at 12,000 *g* to precipitate cellular debris. Supernatants containing plasmid DNA were transferred to fresh tubes and the DNA was precipitated by the addition of two volumes of 100% ethanol and 1/10 volume of 3 M sodium acetate pH 5.2. Solutions were mixed to ensure homogeneity and centrifuged at 12,000 *g* for 15 min at 4 ºC. The supernatants were removed by aspiration, taking care not to disturb the DNA pellet. DNA pellets were washed with 70% ethanol and centrifuged again at 12,000 *g* for 5 min. The final ethanol wash was removed and solutions were dried using a centrifugal vacuum apparatus for 10 min at room temperature to remove any remaining ethanol. Pellets containing plasmid DNA were resuspended in 30 µL of 50 mM Tris•HCl (pH 8.5).

3.2.3.7 DNA Analytical Restriction Digestion and Fragment Analysis

Plasmid DNA obtained from this preparation was subjected to enzymatic digestion with *Nde*I and *Hind*III. A typical reaction mixture contained 15 µL of DNA, 2 µL of 10X NEB reaction buffer #2, 0.1 µL of each enzyme (*Nde*I and *Hind*III), 0.2 µL of 100X BSA and 2.6 µL of sterile water. The reactions were incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis. Samples were diluted 5X in DNA loading dye and electrophoresed for 30 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide (gel concentration).

Electrophoretic results as visualized on agarose gels were used to confirm the presence of the DNA insert of the appropriate length. Colonies determined to contain the

vector and successfully ligated PCR product were grown overnight at 37 ºC with 200 RPM shaking in LB media supplemented with 50 µg/mL kanamycin. After overnight incubation, bacteria were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the bacterial pellet using the Qiagen miniprep kit, following the manufacturer's protocols with the following adjustments to the protocol: the optional 500 µL PB wash was always performed to remove additional contaminants, and elution buffer EB was heated to 65 ºC before DNA elution. The purified plasmid was sequenced to confirm the identity of the PCR product.

PHRED sequences obtained from the DNA facility were aligned against the gene sequences previously identified for the target genes though gene cluster sequencing. Both forward and reverse sequences were compared to the originally obtained gene sequences to determine that the sequences were correct with no mutations.

3.2.3.8 Preparation of Plasmids for S. lividans TK-24 Heterologous Expression

Once the DNA sequence was confirmed to be the desired glycosyltransferase, the gene was digested with the appropriate enzymes, gel purified using the Qiagen gel purification kit following the manufacturer's instructions with adjustments as listed previously. Expression vectors (pWHM3-ermE*, pWHM3-actIp, pWHM467-ermE*, or pWHM467-actIp) were prepared by enzymatic digestion with the *Xba*I and *Hind*III enzymes overnight at 37 ºC, followed by the addition of 0.2 µL (2 units) of calf intestinal alkaline phosphatase (CIP). After CIP addition, the reaction was incubated for an hour at 37 ºC before purification. To purify digested DNA, the Qiagen gel purification kit was used according to manufacturer's instructions for purifying PCR products.

Both vector and gene insert were quantified using a Nanodrop. Ligations were performed at room temperature for 1-2 h or overnight at 4 ºC, with a typical reaction mixture consisting of 90 fmol ends of the gene to be inserted, 30 fmol ends of vector pWHM3, 2 µL of 5X T4 ligase buffer, 0.5 µL of T4 ligase and sterile water to a final volume of 10 µL. An aliquot (5 µL) of the ligation reaction was transferred to 50 µL of chemically competent *E. coli* DH5α and incubated on ice for 30 min. The transformation procedure is identical to that previously listed, with the following adjustments. Ampicillin was the selective antibiotic to be used, thus the transformation was not incubated for 30-45 min after the heat shock, instead, it was plated directly onto LB-agar plates supplemented with 100 µg/mL ampicillin (amp100).

After overnight incubation at 37 ºC, colonies appearing on the selection plates were screened for successful ligation using the alkaline lysis procedure and enzymatic digestion according to the method described above with the following adjustments. The enzymes used for digestion were *Xba*I and *Hind*III. Positive colonies were again amplified and gene-containing plasmids purified for transformation into *S. lividans* TK-24.

3.2.4 PREPARATION OF S. LIVIDANS TK-24 PROTOPLASTS

A starter culture of *S. lividans* was grown in YEME media at 30 ºC with 200 RPM shaking. To inoculate the media, a square of agar was cut from an R1R2 agar plate of *S. lividans* with a red-hot steel needle and introduced to 10-20 mL media in a glass culture tube with sterile glass beads. This culture was grown until mycelia were clearly seen in the media. This small-scale culture was transferred to 50-100 mL of YEME media with 0.5% glycine, and allowed to grow for 48 h at 30 ºC, with 200 RPM shaking. After two

days of growth, very little mycelia was seen in the culture, and the media was somewhat cloudy.

Protoplast buffer (20 mL) (12.875% sucrose, 0.25% $MgCl_2$, 0.03% K_2SO_4, 0.25% trace elements, 0.0625% KH_2PO_4, 4.6% $CaCl_2$, 1.25% 0.25 M TES buffer pH 7.2) and 1 mg/mL lysozyme were added to the solution and mixed gently. The solution was incubated for 30 min at 30 ºC without shaking. After this time, the solution was mixed by swirling gently, and again incubated at 30 ºC without shaking for an additional 15-20 min. The culture was removed from the incubator and filtered in a sterile environment through a syringe packed with sterilized cotton to remove the mycelia from the culture. The filtrate containing protoplasts was centrifuged at 5000 *g* for 30 min. In a sterile environment, the supernatant was decanted, and the pellet containing protoplasts was resuspended in protoplast buffer and divided into 200 µL aliquots. These aliquots were frozen slowly at -80 ºC.

3.2.5 PROTOPLAST TRANSFORMATION OF S. LIVIDANS TK-24

Aliquots (200 µL) containing protoplasted *S. lividans* were removed from the -80 ºC freezer and thawed under hot running water. After thawing, the protoplasts were pelleted at 6700 *g* for 30 s and the supernatant was decanted. The protoplast pellets were resuspended in the remaining protoplast buffer (50-100 µL). The plasmid to be transformed (1 µg) was added to the protoplasts *in situ*, and allowed to equilibrate for several min. To this mixture was added 250 µL of a sterile PEG-T solution (25% polyethylene glycol (PEG) 1450, 75% T-Buffer (2.5% sucrose, 0.25% $MgCl_2$, 1.47% $CaCl_2$, 0.04% K_2SO_4, 0.2% trace elements, 0.58% maleic acid, pH 8.0 with 1 M Tris base, 0.1% KH_2PO_4, 7.36% $CaCl_2$, 2% 0.25 M TES buffer pH 7.2). T-buffer (1 mL) was added

to each of the transformations and 250 μL of each transformation was aliquotted onto five freshly made R1R2 agar plates. The mixture was tapped to distribute evenly on each plate and plates were incubated overnight at 30 ºC. After overnight recovery, antibiotic (generally thiostrepton 50 mg/mL in DMSO) mixed with 10.3% sucrose solution was overlaid onto each plate. Positive transformants were selected after 5 days of incubation at 30 ºC. Distinctive white and red colonies appearing on R1R2 plates were selected and plated onto fresh R1R2 agar supplemented with the appropriate selective antibiotic and incubated at 30 ºC until plates were confluent (5-7 days).

3.2.6 SDS-PAGE

The relative molecular mass and purity of enzyme samples were determined using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli.[273] Proteins were electrophoresed using a discontinuous buffer system 25 mM Tris•HCl, 192 mM glycine and 0.1% SDS (pH 8.3).[273] Gels were 75 mm, and 12% acrylamide was used for the separatory gel, while the stacking gel was 4% acrylamide. Protein samples were diluted 2X with SDS-Loading dye (62.5 mM Tris•HCl buffer pH 6.8, containing 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.0025% bromophenol blue, and heated to 100 ºC for 5 min. Samples were centrifuged briefly to return samples to the bottom of the tubes. The protein gels were electrophoresed using a setting of 30 mA per gel, allowing the gel to run until the loading dye ran off the bottom of the gel. Gels were removed from glass plates, and stacking gel removed. A corner of each gel was also cut to preserve orientation of the gel samples. Gels were stained with Coomassie blue (2.5 g/L of Coomassie Brilliant Blue G-250 in acetic acid:water:methanol (1:4:5) by volume) until gels were dark blue. [278] Gels were removed

from staining solution, rinsed in ddH_2O, and introduced to destaining solution (ethanol:acetic acid:water in a ratio of 4:5:41 by volume). [278] Protein concentrations were determined according to Bradford using BSA as the standard, and the dye reagent from BioRad. [272]

3.2.7 WESTERN BLOT PROCEDURE

Western blotting was used to detect the presence of polyhistidine tags on heterologously expressed and purified proteins. [279] SDS-PAGE of elution fractions from Ni-NTA chromatography procedures was performed in duplicate for each protein. The gels were processed as detailed previously up to the staining step, at which point they were assembled into the transfer cassettes. The gels were transblotted onto the nitrocellulose membrane at 100 V for 1 h, using a running buffer of 25 mM Tris base, 192 mM glycine, 20% v/v methanol (pH 8.3). Cassette assembly and electrophoresis were carried out according to the directions provided by BioRad. After transblotting electrophoresis, the membrane was carefully removed from the transfer cassette and the presence of the prestained marker on the blot was visually assessed to ensure transfer was complete. The membrane was transferred to a small plastic box and immersed in 5% non-fat dry milk in Tris buffered saline and Tween (TBST) (20 mM Tris•HCl, 150 mM NaCl, 0.5% v/v Tween® 20, pH 7.5). The membrane was blocked in this solution for 1 h, with agitation, overnight at 4 ºC or at room temperature for 1 h. After which, the solution was removed and the membrane was rinsed thrice in TBST to remove excess blocking solution. The membrane was immersed in primary antibody solution (15 mL TBST, 1:30,000 dilution of monoclonal anti-polyhistidine clone His-1, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h.

After incubation, the solution was removed from the membrane and the membrane was washed thrice with TBST to remove any excess primary antibody solution remaining on the membrane. The membrane was incubated in secondary antibody solution (15 mL TBST, 1:30,000 dilution of anti-mouse IgG [Fc specific] alkaline phosphatase conjugate, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was removed from the membrane and the membrane was washed thrice with TBST to remove any excess secondary antibody solution remaining on the membrane. To visualize proteins, the membrane was immersed in 5 mL color development solution (100 mM Tris•HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) with the addition of 66 µL NBT and 33 µL BCIP. The membrane was incubated in the dark for 2-5 min or until purple bands appeared on the blot. Development was stopped by rinsing the membrane with ddH_2O. Membranes were dried and scanned within 24 h to preserve coloration.

3.2.8 KIJANIMICIN PRODUCTION

To obtain kijanimicin, kijanosyl-kijanolide and kijanolide, *A. kijaniata* was cultured by the method of Waitz and extracted with ethyl acetate followed by fractionation on silica gel column.[243,68] *A. kijaniata* mycelial glycerol stocks were grown in seed media (2.4% potato starch, 0.5% tryptose, 0.3% beef extract, 0.2% $CaCO_3$, 0.1% glucose, with the solution pH adjusted to 7.2 prior to sterilization). This culture was grown at 30 ºC with 200 RPM shaking for 48 h. Cultures in seed media (10-20 mL) were used to inoculate 1 L of vegetative media (0.5% yeast extract, 1% dextrose, 2% soluble starch, 0.5% NZ-amine A, 0.4% $CaCO_3$, 0.024% $CoCl_2 \cdot 6H_2O$). This culture was allowed

to grow for 5-7 days at 30 ºC with 250 RPM shaking. After this time, cultures were a deep forest green, due to high mycelial growth.

3.2.9 KIJANIMICIN EXTRACTION

To obtain kijanimicins A, B, C and other minor products such as ECK-1, the media containing tetronolide products was separated from *A. kijaniata* mycelia by centrifugation at 6700 *g* for 20 min at 4 ºC. The supernatant was decanted and filtered to remove any unpelleted mycelia. The pellets were combined and washed with sterile water to remove any remaining broth. The mycelial pellets were washed with 10.3% sucrose solution twice and retained at -20 ºC for use as future inoculum. The media was extracted with 2 volumes of ethyl acetate and the solvent was removed through rotary evaporation.

3.2.10 KIJANIMICIN PURIFICATION

STEP 1: SILICA GEL CHROMATOGRAPHY. The crude extract was resuspended in chloroform and loaded onto a silica gel column (4 × 30 cm) prepared in chloroform. The crude mixture was separated using a chloroform-methanol gradient with 2% acetic acid. kijanimicin A, the major product, eluted at 10% methanol + 2% acetic acid, kijanimicin B (**17**) and kijanimicin C (**18**) eluting just prior to kijanimicin A and ECK-1 (**19**) at higher methanol concentrations. Fractions were collected (15 mL) and neutralized with 82.5 µL of 8 N ammonium hydroxide just after elution and assessed for purity by TLC using a mobile phase of 10% methanol in chloroform. TLC plates were developed in vanillin stain (1.5 g vanillin, 3 mL H_2SO_4, in 200 mL methanol), and plates were heated to visualize spots. Kijanimicin compounds appeared as black spots on a pinkish to yellow-

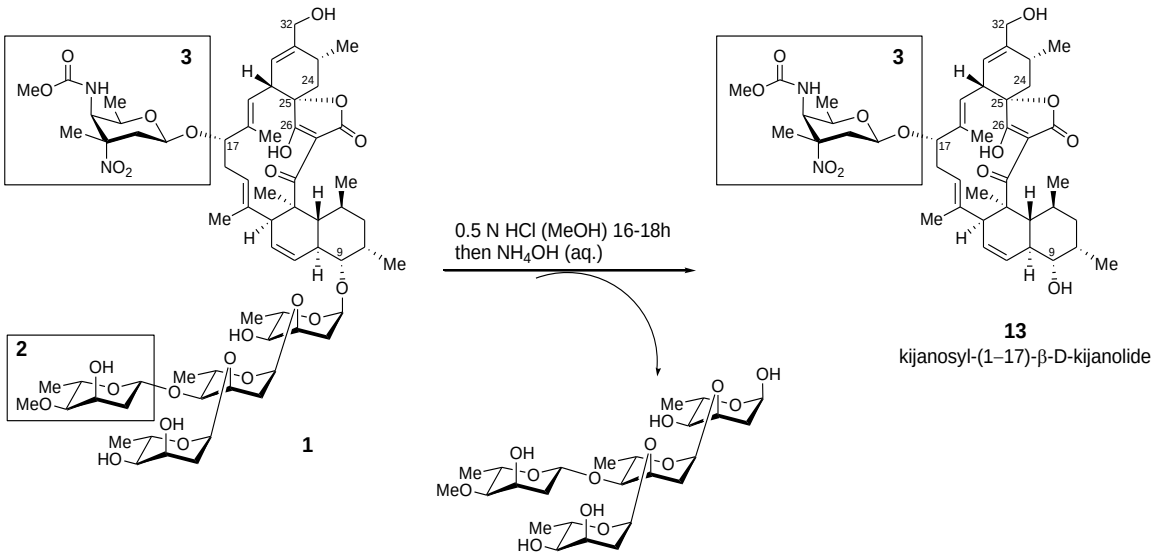
ivory background. Fractions were combined based on TLC purity and were in some cases further purified by HPLC.

STEP 2: HPLC PURIFICATION USING C-18 COLUMN. HPLC purification of kijanimicins was accomplished using a gradient of 20 mM ammonium acetate pH 5.5 as eluent A and acetonitrile as eluent B on a C-18 column. The gradient began in 5% eluent A and remained at 5% for 5 min, after which time, the gradient was increased to 30% B over 10 min, followed by the elution stage, an increase to 70% B over 35 min. After which time, the gradient was increased to 100% B over 2 min, held at 100% B for 5 min and returned to 5% B over 4 min and held at 5% B for 8 min to equilibrate the column. The flow rate was 1 mL/min when using an analytical scale column (250 × 5 mm), and 5 mL/min using a semi-preparatory scale column (250 × 10 mm). The UV detector was set at 240 nm to detect kijanimicin compounds. Under these conditions, kijanimicin A (**1**), the major product, eluted at 33 min, kijanimicin B (**17**) eluted at 41.1 min, and kijanimicin C (**18**) at 47.4 min. The minor product *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) eluted at 27.5 min and ECK-1 (**19**) at 29.8 min. Peaks were collected and submitted for Mass Spectrometry. Dr Hyung Jin Kwon performed the ^{1}H and ^{13}C NMR analysis to further characterize the products.

3.2.11 KIJANIMICIN HYDROLYSIS

To afford *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**), mild acid hydrolysis following the method of Mallams as shown in Scheme 3.2. [245] Briefly, purified kijanimicin A (**1**) (approx. 400 mg, 304 μmol, 1 eq) was introduced to 400 mL of 0.5 N HCl in methanol and stirred for 16-20 h at room temperature. After incubation, the reaction was neutralized with 8 N NH_4OH (to pH 7 as judged by pH paper).

3.2 Kijanimicin Hydrolysis



The resulting crude reaction mixture was dried and resuspended in chloroform. The resulting products were concentrated, and the sample was subjected silica gel chromatography employing a methanol gradient from 1-20% to isolate the *O*-β-D-kijanosyl-(1→17)-kijanolide product **13**. Column fractions observed to contain kij-kij (**13**) were concentrated and subjected to HPLC, following the same procedure described in the previous section. Peaks eluting at 27.5 min were collected from the HPLC, concentrated and submitted to MS to confirm the presence of *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) (calculated m/e $[M-H]^-$ for $C_{42}H_{57}N_2O_{12}$, 781.3912, observed 781.3925). Dr Hyung Jin Kwon performed ^1H and ^{13}C NMR analysis to further characterize product **13**. For the sake of simplicity, protons on the kijanose ring are denoted by "**A**". ^1H NMR (500 MHz, $CDCl_3$) of *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**): 6.01 (1H, ddd, *J*= 10.2, 11-H), 5.49 (1H, s, 21-H), 5.42 (1H, ddd, *J*= 10.0, 4.8, 2.4, 12-H), 5.18 (1H, d, *J*= 9.8,

15-H), 5.13 (1H, d, *J*= 10.9, 19-H), 5.07 (1H, d, *J*= 10.0, **A**-4-N*H*COOCH$_3$), 4.44 (1H, dd, *J*= 9.8, 2.1, **A**-1-H), 4.36 (1H, d, *J*= 9.8, **A**-4-H), 4.21 (2H, s, 22-CH$_2$OH), 4.19 (1H, m, 17-H), 3.70 (3H, s, **A**-4-NHCOO*CH$_3$*), 3.64 (1-H, dd, *J*= 9.5, 6.6, 9-H), 3.60 (1-H, dd, *J*= 10.9, 2.0 20-H), 3.48 (1H, d, *J*= 6.0 13-H), 3.47 (1H, s **A**-5-H), 2.77 (1-H, d, *J*= 15.4, **A**-2-H), 2.66 (1H, dd, *J*= 7.8, 7.4, 23-H), 2.38 (1H, dd, *J*= 14.3, 7.8, 24-H), 2.32 (1H, dd, *J*= 16.2, 9.7, 16-H), 2.25 (1H, br m 16-H), 2.23 (1H br m 8-H), 2.07 (1H, dd, *J*= 10.3, 2.1, 10-H), 1.98 (1H, s 5-H), 1.83 (1H, d, *J*= 14.6, 24-H), 1.62 (3H, s, 4-*CH$_3$*), 1.60 (1H, m, **A**-2-H), 1.60 (1H, s 6-H), 1.57 (3H, s, **A**-3-*CH$_3$*), 1.38 (3H, br s, 18-*CH$_3$*), 1.35 (1H, br s, 14-*CH$_3$*), 1.31 (3H, d, *J*= 7.4, 23-*CH$_3$*), 1.14 (3H, d, *J*= 6.4, **A**-6-H), 1.03 (3H, d, *J*= 7.1, 8-*CH$_3$*), 0.62 (3H, d, *J*= 5.7, 6-*CH$_3$*); ^{13}C NMR (125 MHz, CDCL$_3$): 206.5 (C3), 201.5 (C26), 167.1 (C1), 157.3 (**A**-4-NH*C*OOCH$_3$), 141.4 (C22), 137.1 (C18), 135.8 (C14), 126.6 (C12), 125.4 (C11), 123.4 (C15), 121.5 (C19), 119.3 (C21), 101.9 (C2), 97.1 (**A**-C1), 91.0 (**A**-C3), (C9), 83.2 (C25), 78.5 (C17), 76.1 (C9), 69.0 (**A**-C5), 65.0 (C32), 53.7 (**A**-C4), 53.2 (C13), 52.7 (**A**-4-NHCOO*C*H$_3$), 51.0 (C4), 42.8 (C5), 41.7 (C7), 40.4 (C20), 38.2 (C8), 35.8 (**A**-C2), 35.4 (C24), 34.7 (C8), 31.3 (C23), 31.1 (C6), 31.0 (C16), 22.3 (6-CH$_3$), 20.2 (23-CH$_3$), 17.0 (**A**-C6), 15.1 (4-CH$_3$) 15.0 (18-CH$_3$), 13.7 (14-CH$_3$), 13.0 (8-CH$_3$).

3.2.12 CONSTRUCTION OF THE TDP-L-DIGITOXOSE BIOSYNTHETIC PLASMID

This plasmid was constructed by Dr. Hyung-jin Kwon of our research group. Briefly, a 2.1-kb fragment comprising *kijD4* and *kijD5* was amplified from the cosmid HZ-14 by PCR using the forward primer 5′-TCTTAGAA*TTAATTAA*GAAGCCTTTTGT-AC-3′ with an engineered *Pac*I restriction site (italicized and underlined) and the reverse primer 5′-GCGCCGCC*ATGCAT*TCGAGCAGGACGCGAC-3′ with the engineered *Nsi*I

restriction site (italicized and underlined) and subcloned into the *Pac*I-*Nsi*I sites of pWHM468 to generate p1-86A. A 1.7-kb fragment comprising *kijD10* and *kijD11* was amplified from the cosmid HZ-14 by PCR using the forward primer 5′-CATGACGG*CTGCAG*GTGGGCGCGCGCCATG-3′ with an engineered *Pst*I restriction site and the reverse primer 5′-GCATGACA*AGATCT*CTTGCGTTTGTGCCGA-3′ with the engineered *Bgl*II restriction site (as indicated by italicized and underlined bases) and subcloned into the *Pst*I-*Bgl*II sites of pLitmus28 to generate p1-86C. A 1.8-kb fragment comprising *kijC2* was amplified from the cosmid HZ 4-6 by PCR using the forward primer 5′-CCTCAATG*TTAATTAA*GGACGCCCGTGCCC-3′ with an engineered *Pac*I restriction site, and the reverse primer of 5′-GTGGACACCGCC*ATGCAT*TCTCCTCTT-TCC-3′ with the engineered *Nsi*I site (as indicated by italicized and underlined bases) and subcloned into the *Pac*I-*Nsi*I sites of pWHM468 to generate p1-86B.

A 1.2-kb fragment comprising *kijB1* was amplified from the cosmid HZ 4-6 by PCR using the forward primer 5′-AGCCCTGGT*CTGCAG*ACCGGGGCCGGACAC-3′ with an engineered *Pst*I restriction site and the reverse primer 5′-CTC*GAATTC*G*AGATC-T*CGGGCATCGCCGTC-3’ with the engineered *EcoR*I *and Bgl*II restriction sites (as indicated by italicized and underlined bases), and subcloned into the *Pst*I-*EcoR*I sites of pUC18 to generate p1-86D.

The *Xba*I-*Nsi*I fragment of 1-86A and the *Pst*I-*Bgl*II fragment of 1-86D were subcloned into the corresponding *Xba*I-*Bgl*II sites of pLitmus28, generating p1-94A. The *Xba*I-*Nsi*I fragment of 1-86B and the *Pst*I-*Bgl*II fragment of 1-86C were similarly ligated into the *Xba*I-*Bgl*II sites of pLitmus28, generating pHJK-1-94B. The *Xba*I-*Pac*I fragment containing the actI promoter from pWHM467, the *Pac*I-*Spe*I fragment of 1-94A (*kijD5-kijD4-kijB1*), and the *Xba*I-*Bgl*II fragment of 1-94B (*kijC2-kijD11-kijD10*) were

ligated into the *Xba*I-*Bgl*II restriction sites of 1-107A, which is a pSET152-derivative that harbors an engineered *Bgl*II site in the polycloning site, to yield p1-114A, or pDIG.

3.2.13 EXPRESSION CONSTRUCTS FOR GLYCOSYLTRANSFERASES (IN VIVO)

3.2.13.1 Generation of the kijC3 Expression Construct (In Vitro and In Vivo)

A 1.2-kb fragment comprising *kij*C3 was amplified from the cosmid HZ4-6 by PCR using the forward primer 5′-GTCCTCCCACCGAAGGAGCT*CATATG*CGCG-3′ with an engineered *Nde*I site as indicated above by italicized bases and the reverse primer 5′-TCATGGTCTG*GAATTC*CGGTCGAACGGGCG-3′ with an engineered *EcoR*I site, as indicated by italicized bases. The PCR product was subcloned into the corresponding *Nde*I-*EcoR*I sites of pET28a to generate p1-61A. The *kij*C3 containing insert of pHJK-1-61A was isolated by restriction digest as an *Xba*I-*Hind*III fragment, which includes the ribosomal binding site of pET28a, and ligated into the corresponding restriction sites of pWHM3-ermE* to yield p1-73A for the expression of the *kij*C3 gene.

3.2.13.2 Generation of the kijC4 Expression Construct (In Vitro and In Vivo)

A 1.2-kb fragment comprising *kijC4* was amplified from the cosmid HZ4-6 by PCR using the forward primer 5′-CCCGCCGTCCGCAGAGGTAG*CATATG*CGCA-3′ with an engineered *Nde*I site as indicated by italics, and the reverse primer with an engineered *EcoR*I restriction site 5′-GTGGGAGGAC*GAATTC*CCGGTGGCCCGTCA-3′. The PCR product was subcloned into the corresponding *Nde*I-*EcoR*I sites of pET28a to generate p1-61B. The *kijC4* containing insert of pHJK-1-61B was isolated by restriction digest as a *Xba*I-*Hind*III fragment, which includes the ribosomal binding site

of pET28a, and ligated into the corresponding restriction sites of pWHM3-ermE* to yield p1-73B, for the expression of the *kij*C4 gene.

3.2.13.3 Generation of the kijD9 Expression Construct:

A 1.2-kb fragment comprising *kij*D9 was amplified from the cosmid HZ14 by PCR using the forward primer 5′-GACGGGGGCGCGC*GAATTC*TACTTCCCCTG-3′ with an engineered *EcoR*I site as indicated by italicized bases, and the reverse primer 5′-CGTGCCGTCCTCGG*ATGCAT*GTGCTACCCC-3′, with the engineered *Nsi*I site as indicated by italicized bases. This PCR product was subcloned into the corresponding *EcoR*I-*Nsi*I sites of pLitmus28 to yield p1-77B. The *kij*D9 containing insert of 1-77B was reisolated as *Xba*I-*Nsi*I fragment and ligated into the corresponding *Xba*I-*Hind*III sites of pWHM3-ermE* in conjunction with the *Pst*I-*Hind*III fragment of *kij*D4-*kij*D5 from 1-86A, to generate p1-125B. The insert of p1-77B was recovered as an *Xba*I-*Nsi*I fragment and subcloned into the *Xba*I-*Hind*III sites of pWHM3-ermE* in conjunction with the *Nsi*I-*Hind*III fragment of actIIorf4, to generate p1-125C. Strains containing this plasmid were used as controls for all *in vivo* feeding experiments.

3.2.13.4 Generation of the kijC1 Expression Construct:

The 6-kb *Pst*I fragment from the cosmid HZ4-6 was subcloned into the corresponding *Pst*I site of pLitmus28, to yield p1-115B. Restriction digestion-analysis indicated that an *Xba*I site is present in the polycloning site located upstream to the *kij*C1 gene. Since a *Pst*I site is located 1-kb-upstream of the start codon of *kijC1*, the *kijC1* containing insert of pHJK-1-115B was recovered as a 3-kb *Xba*I-*Sac*I fragment, which also contains truncated *kijB1* upstream and truncated *kijC2* downstream. This fragment

was subcloned into the corresponding sites of pGEM7zf, yielding p1-126A. The *kijC1* containing insert of p1-126A was rescued as an *Xba*I-*Nsi*I fragment and subcloned into the *Xba*I-*Hind*III restriction sites of pWHM3-ermE* in conjunction with the *Nsi*I-*Hind*III fragment of actIIorf4, to generate p2-11C.

3.2.13.5 Generation of the kijA4 Expression Construct:

A 4.7-kb *Nco*I fragment from the cosmid HZ4-6 was ligated into the corresponding *Nco*I site of pLitmus28. The restriction digestion-analysis identified two clones, p1-141A and p1-141B. The *Xba*I site of the polycloning site is located upstream of the *kijA4* gene in p1-141B. It should also be noted that the plasmid p1-141B harbors *kijA3*, in addition to *kijA4*. The *kijA3* and *A4* containing insert in p1-141B was recovered as an *Xba*I-*Hind*III digested fragment and ligated into the corresponding restriction sites of pWHM3-ermE* to generate p2-11B.

3.2.13.6 Generation of the Expression Construct for kijD9, kijC3, kijC4, & kijC1:

The *kijC1* containing insert of pHJK-1-115B was recovered as a 3-kb *Xba*I-*Sac*I fragment, within which the upstream *kijB1* and downstream *kijC2* genes were both truncated, and subcloned into the corresponding restriction sites of pUC18, yielding p1-126C. The *Xba*I-*EcoR*I fragment from p1-126C and the *kijD9* containing *EcoR*I-*Nsi*I fragment from p1-77B were ligated into the *Xba*I-*Nsi*I sites of pLitmus28 to generate p1-128A. The *kijC1-kijD9* containing insert of p1-128A was recovered as an *Xba*I-*Nsi*I fragment and ligated into the corresponding restriction sites of pWHM467, which yielded p1-139A. The 2.5-kb *Xba*I-*EcoR*I fragment containing *kijC3* and *kijC4* from p1-77C was ligated into the corresponding restriction sites of pWHM468 to generate p1-139B, from

which the gene containing insert was isolated as a 2.5-kb *Pac*I-*EcoR*I fragment and ligated into the corresponding restriction sites of pWHM467 to generate p1-142B. The gene-containing insert of p1-139A was isolated as a *Pac*I-*Nsi*I fragment and ligated into p1-142B, previously digested with *Pac*I and *Nsi*I to generate p2-11A.

3.2.14 KIJANIMICIN GLYCOSYLTRANSFERASES C1, C3, C4 AND A4: IN VIVO CONSTRUCTION OF THE TETRASACCHARIDE BRANCHED CHAIN

STEP 1: PREPARATION OF PDIG PROTOPLASTS. The pDIG plasmid containing the entire digitoxose biosynthetic pathway was transformed into *S. lividans* TK-24 protoplasts according to the procedure described previously. After selection on R1R2 plates supplemented with 50 µg/mL apramycin, a plate with confluent mycelial growth, expressing the secondary metabolites (as observed by pigment in agar plates) common to this strain, was used to inoculate a 10-20 mL culture of YEME media supplemented with 50 µg/mL apramycin as described previously (section 3.2.4). The culture was grown at 30 ºC with 250 RPM shaking for 48-72 h, when visible mycelia could be observed in the culture. At this time, 10 to 15 mL of the small-scale culture was used to inoculate 50-100 mL of fresh YEME, supplemented with 50 µg/mL apramycin and 0.5% glycine. This culture was grown at 30 ºC for 3 d with 250 RPM shaking and pDIG protoplasts were prepared from this culture following the procedure described in section 3.2.4.

STEP 2: TRANSFORMATION OF GLYCOSYLTRANSFERASE EXPRESSION PLASMIDS INTO PDIG PROTOPLASTS. A plasmid containing a *kij* glycosyltransferase gene under the control of the ermE* promoter was transformed into *S. lividans* TK-24 pDIG protoplasts according to the procedure described previously (section 3.2.4 and 3.2.5). After selection, a plate with confluent mycelial growth, expressing the secondary metabolites (as seen by visible pigment) common to this strain, was used to inoculate a 20 mL culture of YEME

media supplemented with 50 µg/mL thiostrepton and 50 µg/mL apramycin using the procedure described previously (section 3.2.5). The culture was grown in a sterile culture tube with glass beads at 30 ºC with 250 RPM shaking for 48-72 h, at which point mycelia were observed in the culture.

Step 3: Preparation of pDIG-GT Strains for Feeding Experiments. At this time, the small-scale cultures were used to inoculate fresh R1R2 media (with and without sucrose), supplemented with 50 µg/mL thiostrepton and 50 µg/mL apramycin. Plates were inoculated with 1 mL of culture and grown at 30 ºC for 3-5 d, or until secondary metabolites (pigment) could be observed in the agar.

Step 4: Feeding Experiments Using pDIG-GT Strains. After 3-4 d, the plates were observed to have strong mycelial growth and had turned a reddish-purple color from secreted secondary metabolites. At this stage of growth, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**), kijanimicin A (**1**), kijanimicin B (**17**), or kijanimicin C (**18**) in cold 10.3% sucrose solution (5 µg/mL agar, 200 µg/plate) was spread onto the plates and allowed to dry. The plates were returned to the incubator and grown at 30 ºC for an additional 3-5 d.

Step 5: Harvesting Agar from Feeding Experiments Using pDIG-GT Strains. After the final incubation, the agar was removed from the plastic dishes and transferred to 2 L wide mouthed Erlenmeyer flasks. Ethyl acetate to cover the plates was added to each flask and the agar was extracted overnight. After 16-18 h, the organic extracts were dried over Na_2SO_4. The organic extracts were then transferred to 1L round-bottomed flasks and the solvent was removed by rotary evaporation. The wine-colored residue was resuspended in $CHCl_3$ and transferred to clean glass sample bottles.

Step 6: Feeding Experiment Sample Preparation: Silica Gel Chromatography. The crude mixture was separated using a chloroform-methanol step

gradient in the presence of 2% acetic acid. The column was flushed with CH_3Cl until the first yellow pigment eluted and the gradient started at that time. Kijanimicin A (**1**) eluted at 10% methanol + 2% acetic acid, whereas kijanosyl-kijanolide (kij-kij) (**13)**, kij B (**17**) and kij C (**18**) eluted prior to kij A. Generally, methanol steps included 1%, 2%, 5% and 10% and each step included 2% acetic acid. Fractions were collected (15 mL) until the purple pigment band eluted from the column. Fractions were immediately neutralized with 82.5 µL of 8 N ammonium hydroxide. Aliquots (1 mL) of each fraction were transferred to Eppendorf tubes and the solvent was removed using a speedvac. The concentrated samples were assessed for the presence of kijanimicin compounds by TLC using a mobile phase of 10% methanol in chloroform. TLC plates were developed in vanillin stain (1.5 g vanillin, 3 mL H_2SO_4, in 200 mL methanol), and plates were heated to visualize spots. Kijanimicin compounds appeared as black spots on a pinkish to yellow-ivory background. Fractions were combined based on TLC results and analyzed using HPLC.

STEP 7: HPLC ANALYSIS OF FEEDING EXPERIMENT SAMPLES USING C-18 COLUMN. HPLC analysis of kijanimicins from feeding experiments was accomplished using several different HPLC programs, although all samples were processed using a gradient of 20 mM ammonium acetate pH 5.5 as eluent A and acetonitrile as eluent B on a C-18 column. For analytical separations the flow rate was 1 mL/min, while for preparatory separations the flow rate was 5 mL/min. In all cases, the detector was set to 240 nm. For the initial glycosyltransferase activity test, the gradient began in 10% eluent A and remained at 10% for 10 min, after which time, the gradient was increased to 95% B over 20 min. After which time, the gradient was held at 95% B for 10 min and returned to 10% B to equilibrate the column. Under these conditions, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) eluted at 22 min *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-kijanolide (1-

SC) (**14)** at 22.5 min, and *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-*O*-α-L-digitoxosyl-(1→3)-kijanolide (2-SC) (**15)** at approximately 22.5 min. To assess the production of **15**, and afford resolution between **13**, **14** and **15** the following program was used. The gradient began in 5% eluent A and remained at 5% for 10 min, after which time, the gradient was increased to 30% B over 20 min, followed by the elution stage, an increase to 40% B over 30 min. After which time, the gradient was increased to 100% B over 2 min, held at 100% B for 5 min and returned to 5% B over 4 min and held at 5% B for 8 min to equilibrate the column. Under these conditions, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) eluted at 48 min *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-kijanolide (1-SC) (**14)** at 52 min, and *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-*O*-α-L-digitoxosyl-(1→3)-kijanolide (2-SC) (**15)** at approximately 53 min. For the glycosyltransferase activity test using (**18**) as the substrate, the gradient began in 10% eluent A and remained at 10% for 10 min, after which time, the gradient was increased to 80% B over 20 min. After which time, the gradient was held at 95% B for 10 min and returned to 10% B to equilibrate the column. Under these conditions, kijanimicin C (**18**) eluted at 28.5 min, and 32-deoxy-*O*-β-D-kijanosyl-(1→17)-kijanolide (**20**) eluted at 29 min. All other experiments employed a gradient that began in 5% eluent A and remained at 5% for 5 min, after which time, the gradient was increased to 30% B over 10 min, followed by the elution stage, an increase to 70% B over 35 min. After which time, the gradient was increased to 100% B over 2 min, held at 100% B for 5 min and returned to 5% B over 4 min and held at 5% B for 8 min to equilibrate the column. Under these conditions using an analytical scale column, kijanimicin A (**1**), the major product, eluted at 33 min, kijanimicin B (**17**) eluted at 41.1 min, kijanimicin C (**18**) at 47.4 min, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) at 27.5 min and ECK-1 (**19**) at 29.8 min.

3.2.15 KIJANIMICIN GLYCOSYLTRANSFERASE C4: IN VIVO GENERATION OF COMPOUND 15

Using the previously constructed *S. lividans* strain containing the pDIG and KijC4 plasmids, the following feeding experiments were carried out.

STEP 1: FEEDING EXPERIMENTS USING PDIG-GTC4 STRAINS. After 3-4 d, the plates were observed to have strong mycelial growth and had turned a reddish-purple color from secreted secondary metabolites. At this stage of growth, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) in cold 10.3% sucrose solution (5 µg/mL agar, 200 µg/plate) was spread onto the plates and allowed to dry. The plates were returned to the incubator and grown at 30 ºC for an additional 3-5 d.

STEP 2: HARVESTING AGAR FROM FEEDING EXPERIMENTS USING PDIG-GT STRAINS. After the final incubation, the agar was removed from plastic dishes and transferred to 2 L wide mouthed Erlenmeyer flasks. Ethyl acetate to cover the plates was added to each flask and the agar was extracted overnight. After 16-18 h, the organic extracts were dried over Na_2SO_4, and the solvent was removed by rotary evaporation. The wine-colored residue was resuspended in $CHCl_3$ and transferred to clean glass sample bottles.

STEP 3: FEEDING EXPERIMENT SAMPLE PREPARATION: SILICA GEL CHROMATOGRAPHY. The crude mixture was separated using a chloroform-methanol step gradient in the presence of 2% acetic acid. Generally, methanol steps included 1%, 2%, 5% and 10% and each step included 2% acetic acid. *O*-β-D-kijanosyl-(1→17)-kijanolide (kij-kij) (**13**), *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-kijanolide (1-SC) (**14**), *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-*O*-α-L-digitoxosyl-(1→3)-kijanolide (2-SC) (**15**), eluted within the first 10-20 fractions (1-2% methanol). Fractions were collected (15 mL) until the purple pigment band eluted from the column. Fractions were immediately neutralized with 82.5 µL of 8 N ammonium hydroxide. Aliquots (1 mL) of

each fraction were transferred to Eppendorf tubes and the solvent was removed using a speedvac. The concentrated samples were assessed for the presence of kijanimicin compounds by TLC using a mobile phase of 1:2 hexanes: ethyl acetate with 2% acetic acid. TLC plates were subjected to development twice in the same solvent system before visualization. TLC plates were developed in vanillin stain (1.5 g vanillin, 3 mL H_2SO_4, in 200 mL methanol), and plates were heated to visualize spots. Kijanimicin compounds appeared as black spots on a pinkish to yellow-ivory background. Compounds **13**, **14** and **15** had approximate R_f values of 0.31, 0.11 and 0.06, respectively. Fractions were combined based on TLC results and the organic solvent was removed under vacuum. The remaining aqueous solvent was removed using the speedvac until samples were dry. The dry samples were resuspended in methanol and further purified using HPLC.

STEP 4: HPLC PURIFICATION OF COMPOUND 15 USING C-18 COLUMN. HPLC purification of (**15**) from feeding experiments was accomplished using several different HPLC programs, although all samples were processed using a gradient of 20 mM ammonium acetate pH 5.5 as eluent A and acetonitrile as eluent B on a semi-preparative scale C-18 column, with a flow rate of 5 mL/min and the detector was set to 240 nm. To afford resolution between **13**, **14** and **15**, the gradient began in 5% eluent A and remained at 5% for 10 min, after which time, the gradient was increased to 30% B over 20 min, followed by the elution stage, an increase to 40% B over 30 min. After which time, the gradient was increased to 100% B over 2 min, held at 100% B for 5 min and returned to 5% B over 4 min and held at 5% B for 8 min to equilibrate the column. Under these conditions, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) at 42 min *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-kijanolide (1-SC) (**14)** at 46 min, *O*-β-D-kijanosyl-(1→17)-*O*-α-L-digitoxosyl-(1→9)-*O*-α-L-digitoxosyl-(1→3)-kijanolide (2-SC) (**15)** at approximately 50 min.

3.2.16 KIJANIMICIN SUSCEPTIBILITY OF S. LIVIDANS

S. lividans TK-24 was cultured on agar plates, prepared using R1R2 media and 40 μg of each compound, kijanimicin A (**1**), kijanimicin B (**17**), kijanimicin C (**18**), and *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) were applied in a Φ of 0.1 cm. The plate was incubated at 30 ºC for 2-3 d, and growth inhibition radius for each compound was measured.

3.2.17 EXPRESSION CONSTRUCTS FOR GLYCOSYLTRANSFERASES (IN VITRO)

Generation of the *kijC3* and *kijC4* expression constructs (1-73A and 1-73B respectively) were described in sections 3.2.13.1 and 3.2.13.2.

3.2.17.1 Generation of the KijD9 In Vitro Expression Construct:

A 1.2-kb fragment comprising *kij*D9 was amplified from the cosmid HZ-14 by PCR using the forward primer 5′-GATCGAT*CATATG*CGGGTTCTGTTCACC-3′ with an engineered *Nde*I site as indicated by italicized bases, and the reverse primer 5′-TATT*AAGCTT*CTACCCCCGGACCAGAGC-3′ with the engineered *Hind*III site as indicated by italicized bases. Protocols for PCR, cloning, digestion and transformation were followed as described above. The plasmid generated by this protocol is *pWHM3(ermE*)-kijd9*, which produces an *N*-terminal His_6 tagged KijD9 protein.

3.2.17.2 Generation of the KijA4 In Vitro Expression Construct:

A 1.2-kb fragment comprising *kij*A4 was amplified from the cosmid HZ-14 by PCR using the forward primer 5′-CACGATC*CATATG*CTGTTGCGAATCCTGTTCAC-CG-3′ with an engineered *Nde*I site as indicated by italicized bases, and the reverse primer 5′-CATG*AAGCTT*GGCCGGTGCGGTCAC-3′ with the engineered *Hind*III site as indicated by italicized bases. Protocols for PCR, cloning, digestion and transformation were followed as described above. The plasmid generated by this protocol is *pWHM3(ermE*)-kija4*, which produces an *N*-terminal His_6 tagged KijA4 protein.

3.2.17.3 Generation of the KijC1 In Vitro Expression Construct:

A 1.2-kb fragment comprising *kijC1* was amplified from the cosmid HZ-14 by PCR using the forward primer 5′-CTCCGTT*CATATG*CGCGTTCTGTTCACCCTG-3′ with an engineered *Nde*I site as indicated by italicized bases, and the reverse primer 5′-GAACAAGCTTTCAGACCAGCCGCTCCAG-3′ with the engineered *Hind*III site as indicated by italicized bases. Protocols for PCR, cloning, digestion and transformation were followed as described above. The plasmid generated by this protocol is *pWHM467(ermE*)-kijc1*, which produces an *N*-terminal His_6 tagged KijC1 protein.

3.2.17.4 General Procedure for the Growth of Cells Expressing Kijanimicin Glycosyltransferases

A plasmid containing a *kij* glycosyltransferase gene under the control of the ermE* promoter was transformed into *S. lividans* TK-24 protoplasts according to the procedure described previously. After selection, a plate with confluent mycelial growth, expressing the secondary metabolites (as seen by visible pigment) common to this strain, was used to inoculate a 20 mL culture of YEME media supplemented with 50 μg/mL

thiostrepton as previously described in section 3.2.4. The culture was grown in a sterile culture tube with glass beads at 30 ºC with 250 RPM shaking for 48-72 h, when mycelia was observed in the culture. At this time, 10 to 15 mL of the small-scale culture was used to inoculate 1 L of fresh YEME, supplemented with 50 µg/mL thiostrepton. This culture was grown at 30 ºC for 3 days with 250 RPM shaking.

After three days, the culture was observed to have strong mycelial growth and had turned a pink/purple color from excreted secondary metabolites. At this stage of growth, the mycelial pellet was harvested by centrifugation at 6700 *g*, 4 ºC for 20 min. After the initial centrifugation, the pellet was washed with water and re-pelleted under the same conditions to remove any remaining media. The mycelia were stored at -80 ºC until protein extraction or processed immediately.

3.2.17.5 Purification of the Kijanimicin Glycosyltransferase Proteins from S. lividans.

STEP 1: CRUDE EXTRACT PREPARATION. The mycelia were resuspended in 20 mL of Lysis Buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole with 10% glycerol, pH 8.0) and 1 mg/mL lysozyme was added to the mixture. The solution was stirred every 15 min for 1 h and sonicated for a total pulse time of 3 min, using a program of 10 s bursts with 30 s cooling intervals to complete the lysis of the bacterial cell wall. After sonication, the solution was clarified by centrifugation at 4 ºC for 20 min at 10,000 *g*.

STEP 2: AMMONIUM SULFATE PRECIPITATION. The supernatant was decanted into a chilled 40 mL centrifuge tube on ice and a small stir bar was added. Ammonium sulfate (to 70%) was added slowly over 30 min while the supernatant was stirred slowly on ice. After the ammonium sulfate was added, the solution was transferred to an orbitron mixing device at 4 ºC and allowed to mix for 30 min to 1 h. After the solution was mixed,

the ammonium sulfate precipitated solution was centrifuged at 18,000 *g* for 30 min at 4 ºC to isolate the proteins. The supernatant was carefully decanted and the pellet containing the protein of interest was resuspended in 5-10 mL of Lysis buffer II, and dialyzed against 3 L of Dialysis buffer IIa (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole with 10% glycerol, pH 8.0), at 4 ºC, for 3 h with changes of 1 L at the end of each h.

STEP 3: NI-NTA CHROMATOGRAPHY. After dialysis, the protein mixture was introduced to 2 mL of Ni-NTA resin prepared by washing three times in Lysis buffer II, and bound overnight at 4 ºC on the orbitron mixer. After Ni-NTA binding, the slurry was introduced into a 1 × 25 cm open column and allowed to flow through by gravity. The column was washed twice with 25 mL of Wash Buffer II and the protein was eluted with 12 mL of Elution Buffer II. Elution fractions (1 mL) were collected and analyzed by SDS-PAGE and Western Blot. Fractions that appeared to contain the protein of interest as detected by Coomassie staining were combined, dialyzed against 3 L of Dialysis buffer II, at 4 ºC, for 3 h with changes of 1 L at the end of each h. After dialysis, the protein was concentrated under N_2 gas at 4 ºC using an Amicon stirred cell system equipped with a YM-10 membrane. The protein was dispensed into aliquots, flash frozen in liquid N_2 and stored at -80 ºC until use.

3.2.17.6 Purification of KijA4 and KijC1 from E. coli.

STEP 1: GROWTH OF E. COLI CELLS. An overnight culture of *E. coli* BL21(DE3)-pET-*kijA4* or *E. coli* BL21 Rosetta II (DE3)-pET-*kijC1* was grown in LB medium supplemented with kanamycin (50 µg/mL) (and chloramphenicol (35 µg/mL) for *kijC1*) at 37 °C. An aliquot (2 mL) of overnight culture was used to inoculate 1 L of LB

supplemented with kanamycin (50 μg/mL). To express *kijC1*, typically, 3 L of culture was incubated at 30 °C. The culture was allowed to grow for an additional 18 h at 30 °C, at which point the cells were harvested by centrifugation at 4,500 *g* for 15 min at 4 °C, and stored at -80 °C. To express *kijA4*, typically, 3 L of culture was incubated at 30 °C with 250 RPM shaking until the OD_{600} of the culture reached 0.3, at which point the temperature was reduced to 24 ºC. The cultures were allowed to grow for an additional hour and the OD_{600} reached 0.4. The cultures were induced with IPTG to a final concentration of 50 μM, and allowed to grow for an additional 15 h at 18 or 22 ºC. The cells were harvested by centrifugation at 4,500 *g* for 20 min at 4 ºC and stored at –80 ºC until protein purification.

STEP 2: CRUDE EXTRACT PREPARATION. The cells were resuspended in 20 mL of Lysis Buffer II and 1 mg/mL lysozyme was added to the mixture. The solution was stirred every 15 min for 1 h and sonicated for a total pulse time of 2 min, using a program of 10 s bursts with 30 s cooling intervals to complete the lysis of the bacterial cell wall. After sonication, the solution was clarified by centrifugation at 4 ºC for 20 min at 10,000 *g*.

STEP 3: NI-NTA CHROMATOGRAPHY. The supernatant was introduced to 2 mL of Ni-NTA resin prepared by washing three times in Lysis buffer II, and bound overnight at 4 ºC on the orbitron mixer. After Ni-NTA binding, the slurry was introduced into a 1 × 25 cm open column and allowed to flow through by gravity. The column was washed twice with 25 mL of Wash Buffer II and the protein was eluted with 12 mL of Elution Buffer II. Elution fractions (1 mL) were collected and analyzed by SDS-PAGE and Western Blot.

Fractions that appeared to contain the protein of interest as detected by Coomassie staining were combined, dialyzed against 3 L of Dialysis buffer II, at 4 ºC, for 3 h with

changes of 1 L at the end of each h. After dialysis, the protein was concentrated under N_2 gas at 4 ºC using an Amicon stirred cell system equipped with a YM-10 membrane. The protein was dispensed into aliquots, flash frozen in liquid N_2 and stored at -80 ºC until use. Culture and induction conditions were optimized for each enzyme to obtain maximum amounts of soluble protein. The gene products were purified by Ni-NTA affinity chromatography to near homogeneity.

3.2.18 KIJANIMICIN GLYCOSYLTRANSFERASES C1, C3, C4 AND A4: IN VITRO CONSTRUCTION OF THE TETRASACCHARIDE BRANCHED CHAIN

Kijanimicin glycosyltransferases KijC4, KijC3, KijC1, KijA4 and KijD9 were purified from *S. lividans* in the previous section were utilized in activity assays. Kijanolide, obtained through mild acid hydrolysis as described in section 3.2.11 was also employed. [245] TDP-L-digitoxose was biosynthesized according to procedures described in Chapter 2. Each glycosyltransferase protein was incubated with the crude TDP-L-digitoxose and kijanolide substrates to generate the product *in vitro* and reactions were monitored by TLC and analytical HPLC. Reactions were carried out in triplicate and conditions were varied in each trial to test the glycosyltransferase activity.

3.2.18.1 ATTEMPT I

In this set of reactions, the pH varied between 7.5 to 9. Reactions consisted of the following: 0.1 mM crude TDP-L-digitoxose (**10**), 100 µM KijC4,10 mM $MgCl_2$, 1.2 mM kij-kij (in MeOH) (**13**), 1 mg/mL BSA, and 50 mM Tris (pH 7.5-9.0) in a final reaction volume of 180 µL. Aliquots (30 µL) were removed from the reaction at various time points (0 h, 30 min, 1 h, 2 h, 4 h, overnight) and flash frozen. Samples were thawed and proteins were removed using a YM-10 centrifugal filter (20 min 14,000 *g*, 4 ºC). The

flow through was retained and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.2.18.2 ATTEMPT II

In this set of reactions, TDP-L-digitoxose was generated *in situ* using the 4-ketoreductase from the mycarose pathway, TylC2, which was previously purified by Dr. Haruko Takahashi.[306] Reactions consisted of the following: 1 mM TDP-4-keto-2,6-dideoxy-D-glucose, 10 nM KijD11, 10μM TylC2, 100 μM KijC4, 10 mM $MgCl_2$, 1.2 mM kij-kij (**13**) (in MeOH), 1 mg/mL BSA, 1.4 mM NADPH and 50 mM Tris (pH 9.0) in a final reaction volume of 180 μL. Aliquots (30 μL) were removed from the reaction at various time points (0 h, 30 min, 1 h, 2 h, 4 h, overnight) and flash frozen. Samples were thawed and proteins were removed using a YM-10 centrifugal filter (20 min 14,000 *g*, 4 ºC). The flow through was retained and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.2.18.3 ATTEMPT III

In this set of reactions, TDP-L-digitoxose was generated *in situ* using the 4-ketoreductase from the digitoxose pathway, KijC2. Reactions consisted of the following: 1 mM TDP-4-keto-2,6-dideoxy-D-glucose, 10 nM KijD11, 10μM KijC2, 100 μM KijC4 (or other kijanimicin GT), 10 mM $MgCl_2$, 1.2 mM kij-kij (**13**) (in MeOH), 1 mg/mL BSA, 10% glycerol, 1.4 mM NADH and 50 mM Tris (pH 9.0) in a final reaction volume of 200 μL. Aliquots (40 μL) were removed from the reaction at various time points (0 h, 2 h, 4 h, 8h, overnight) and flash frozen. Samples were thawed and proteins were removed using a YM-10 centrifugal filter (20 min 14,000 *g*, 4 ºC). The flow through was retained and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.2.18.4 ATTEMPT IV

In this set of reactions, TDP-L-digitoxose was generated *in situ* using the 4-ketoreductase from the digitoxose pathway, KijC2. The substrates were varied as well, using both **13** and **15** for the next reaction steps. Reactions consisted of the following: 2 mM TDP-4-keto-2,6-dideoxy-D-glucose, 20 nM KijD11, 20 µM KijC2, 30 µM KijGT, 10 mM $MgCl_2$, 1.2 mM 2SC (**15**) or 1.2 mM kij-kij (**13**) (in MeOH), 10% glycerol, 1.4 mM NADH and 50 mM Tris (pH 9.0) in a final reaction volume of 200 µL. Aliquots (40 µL) were removed from the reaction at various time points (0 h, 2 h, 4 h, 8h, overnight) and flash frozen. Samples were thawed and proteins were removed using a YM-10 centrifugal filter (20 min 14,000 *g*, 4 ºC). The flow through was retained and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.2.18.5 ATTEMPT V

In this set of reactions, TDP-L-digitoxose was generated *in situ* using the 4-ketoreductase from the digitoxose pathway, KijC2. The substrates were varied as well, using both **13, 14** and **15** for the aglycone acceptor. Reactions consisted of the following: 2 mM TDP-4-keto-2,6-dideoxy-D-glucose, 20 nM KijD11, 20 µM KijC2, 25 µM KijGT, 10 mM $MgCl_2$, 1.2 mM 2SC (**15**) 1.2 mM 1SC (**14**) or 1.2 mM kij-kij (**13**) (in DMSO), 10% glycerol, 2.4 mM NADH and 25 mM Tris (pH 8.0) in a final reaction volume of 200 µL. Aliquots (50 µL) were removed from the reaction at various time points (0 h, 8 h, 16 h and 24) and quenched in 250 µL methanol. Samples were centrifuged, the supernatant transferred to fresh tubes and concentrated using a speedvac. Samples were resuspended in 100 µL methanol and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.2.19 REVERSE GLYCOSYLTRANSFERASE REACTIONS

KijC1, KijA4 and KijC4 were examined for their ability to catalyze the formation of TDP-L-digitoxose or kijanimicin A biosynthetic intermediates from TDP and kijanimicin A. Reaction conditions were as follows #1) 1 mM kijanimicin A (**1**), 10 mM TDP, 10 mM $MgCl_2$, 10 μM GT, 50 mM Tris pH 9 in a final volume of 180 μL. Reaction #2: 1 mM kijanimicin (**1**), 100 μM TDP, 10 mM $MgCl_2$, ~10 μM GT, 10 μM KijC4, 10 μM kij-kij (**13**), 50 mM Tris pH 7.5 in a final volume of 180 μL. Reaction #3: 1 mM kijanimicin (**1**), 1 mM kij-kij (**13**), 100 μM TDP, 10 mM $MgCl_2$, ~10 μM GT, 1 mg/mL BSA in 50 mM Tris pH 7.5 in a final volume of 180 μL. Reaction #4: 10 mM kijanimicin (**1**), 10 mM TDP, 10 mM $MgCl_2$, ~10 μM GT, 1 mg/mL BSA, 50 mM Tris pH 7.5 in a final volume of 100 μL. Total final concentration of methanol was less than 5% in each reaction. All reactions were incubated overnight, 1-3 were incubated at room temperature and reaction 4 was heated to 30 ºC. Reactions were quenched by the addition of 2 volumes of methanol to precipitate proteins. Proteins were removed from the reactions by centrifugation (20 min 16,100 *g*, 4 ºC). Supernatant was retained and subjected to HPLC equipped with a C-18 column, following the HPLC procedure previously described.

3.3 RESULTS AND DISCUSSION

3.3.1 KIJANIMICIN PRODUCTION AND ISOLATION

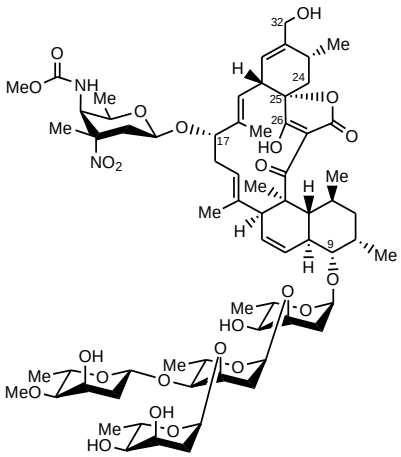
To obtain kijanimicin (**1**) and other kijanolide compounds (**17-19**), *A. kijaniata* was cultured by the method of Waitz and the secreted secondary metabolites were

extracted with ethyl acetate followed by fractionation on silica gel column. [243,68] Cultures of wild-type *A. kijaniata* were grown using the procedure of Waitz et al. with several modifications. [243] The major modification was that the cultures were grown aerobically. The culture media was harvested and the kijanimicin compounds were isolated from the media by extraction with ethyl acetate followed by purification using flash chromatography and HPLC.

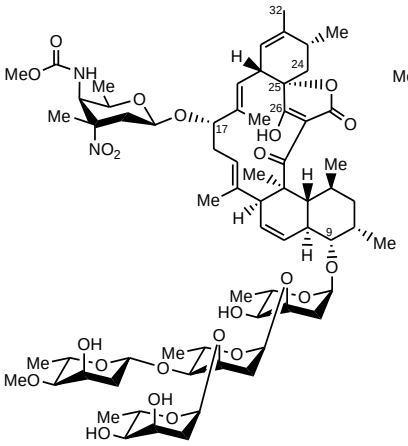
Flash chromatography employed a chloroform-methanol gradient in the presence of 2% acetic acid. Fractions were collected and immediately neutralized with ammonium hydroxide. TLC assessment used a mobile phase of 10% methanol in chloroform, or less frequently hexanes: ethyl acetate (1:2). TLC plates were developed in vanillin stain, and plates were heated to visualize spots. Kijanimicin compounds appeared as black spots on a pinkish to yellow ivory background. Kijanimicin A (**1**), the major product, eluted at 10% methanol + 2% acetic acid, and kijanimicin B (**17**) and C (**18**) eluted just prior to kijanimicin A. Fractions were combined based on purity as assessed by TLC and were in some cases further purified by HPLC. Under the previously reported HPLC conditions, kijanimicin A (**1**), the major product, eluted at 33 min, kijanimicin B (**17**) eluted at 41.1 min, kijanimicin C (**18**) at 47.4 min, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) at 27.5 min, and ECK-1 (**19**) at 29.8 min.

HPLC purified compounds were subjected to analysis by MS and NMR. Kijanimicin A (**1**) was analyzed by positive FT-MALDI-TOF MS to yield m/e 1339.675 (calculated m/e for $[M+Na]^+$ of $C_{67}H_{100}N_2O_{24}$ is 1339.656). The identity of kijanimicin was further verified by ^1H and ^{13}C-NMR analysis, and the spectral data were consistent with previously reported spectra for kijanimicin.[246]

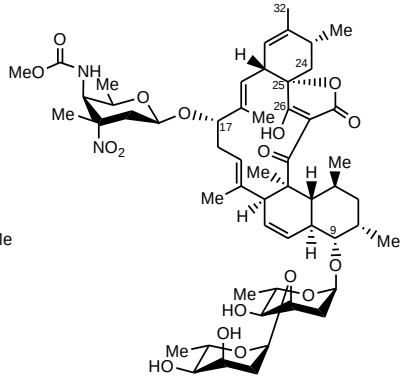
3.4 *A. kijaniata* Secondary Metabolites, Kijanimicins (A, B and C) and Minor Products ECK-1 and *O*-β-D-(1-17)-kijanosyl-kijanolide



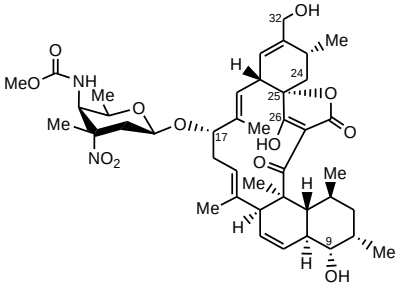
1
Kijanimicin A



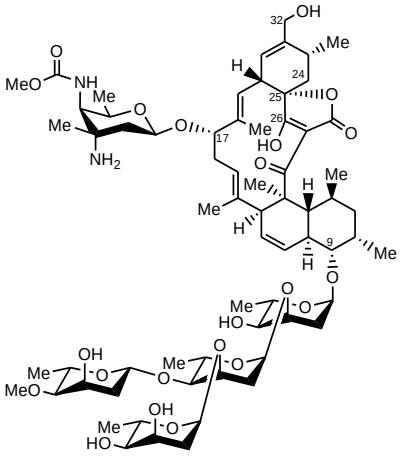
17
Kijanimicin B



18
Kijanimicin C



13
O-β-D-(1-17)-Kijanosyl-Kijanolide



19
ECK-1

High resolution MS (negative FAB mass spectroscopy) for kijanimicins B (**17**) and C (**18**) yielded 1025.5248 (calculated m/e $C_{54}H_{77}N_2O_{17}$, $[M-H]^-$ 1025.5222) and 1299.6645 (calculated m/e for $C_{67}H_{99}N_2O_{23}$, $[M-H]^-$ 1299.6639), respectively. Through NMR analysis, the structures of these two compounds were firmly designated as

kijanimicin B (**17**) and kijanimicin C (**18**). The isolation yields of kijanimicin B (**17**) and kijanimicin C (**18**) were 5 mg and 2 mg per liter of media, respectively.

Other minor compounds were also detected in the HPLC separation. One of these was predicted to be desmethyl kijanimicin B, which gave m/e [M-H]-, 1285 using low resolution FAB mass analysis. Using a semi-preparative HPLC column (and the gradient profile described in section 3.2.10) with a flow rate of 5 mL/min, kijanimicin C (**18**), the proposed desmethyl kijanimicin B, and kijanimicin B (**17**) eluted at 36, 39, and 42 min respectively. However, *O*-β-D-kijanosyl-(1→17)-kijanolide (**16a** or **16b**), which possesses three digitoxose moieties was not found in any of the extraction samples or fractions, although we searched diligently for its presence. This indicated that the three-digitoxose containing compound is not accumulated at detectable levels in *A. kijaniata*.

These results are somewhat surprising if the digitoxose moieties are added sequentially as we originally proposed. If the sugars are indeed added sequentially, we expect to observe compounds carrying 1, 2, 3, and 4 digitoxose sugars. There are three possible interpretations of this data. One possibility is that the export of the 3-sugar compound is somehow impeded, and the cellular export machinery can only export the 2- and 4-sugar compounds. This possibility was quickly discarded, as the addition of one sugar is such a small change of the overall structure of the product. A more likely scenario is that the protein(s) responsible for the latter steps of digitoxose attachment are very fast, much faster than the protein(s) responsible for the initial 2-sugar attachment. Another possibility is that there is only one protein responsible for the last two sugar transfers and there is little protein-substrate dissociation between the successive transfers of the third and fourth sugar. In a sequential transfer scenario catalyzed by a single enzyme, there would be no need for the growing natural product to dissociate from the

active site, thus the addition of the third sugar would be followed immediately by the addition of the fourth and final digitoxose sugar.

To differentiate between the latter two possibilities and to determine the catalytic roles of these glycosyltransferases in sugar attachments, glycosyltransferase assays were carried out to determine the order of attachment of the sugars and which protein is responsible each specific transfer. To carry out these assays, both a source of TDP-L-digitoxose and the kijanolide aglycone are required. In the previous Chapter, the TDP-L-digitoxose biosynthetic pathway was elucidated and thus was applied to prepare the TDP-sugar donor *in vivo* and *in vitro*. A synthetic approach to obtain kijanolide was ruled out due to the complexity of the structure. Thus, we chose to use chemical hydrolysis of the kijanimicin product obtained through bacterial production to prepare the desired kijanolide and its derivatives.

3.3.2 KIJANIMICIN AGLYCONE GENERATION AND PURIFICATION

To generate *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) for use in the glycosylation experiments, we obtained kijanimicin A (**1**) in good yield and removed the tetrasaccharide branched chain by mild acid hydrolysis. After neutralization and flash chromatography, *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) was isolated and identified by high resolution using negative FAB MS, which yielded 781.3925 (calculated for m/e $[M-H]^-$ for $C_{42}H_{57}N_2O_{12}$, 781.3912), and further verified by ^1H and ^{13}C-NMR analysis. From this reaction, 140 mg (35% yield) of *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**), was obtained. This aglycone was used in both *in vitro* and *in vivo* biosynthetic assays.

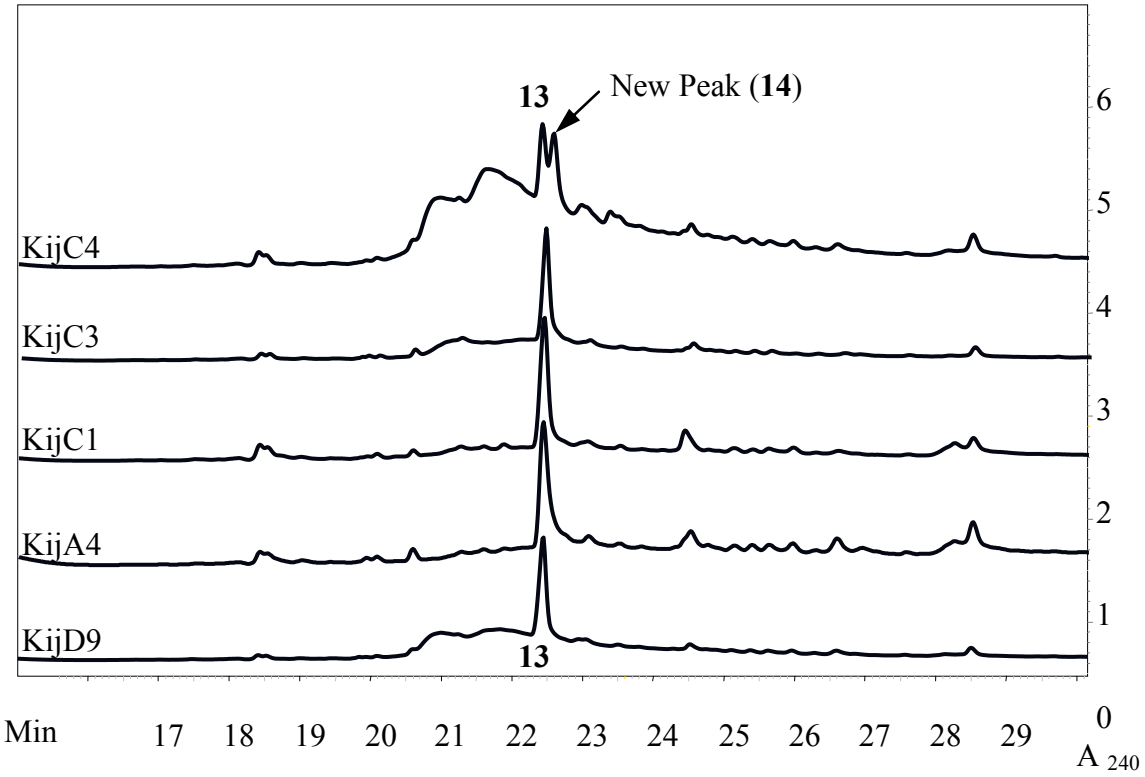
3.3.3 KIJANIMICIN GLYCOSYLTRANSFERASES C1, C3, C4 AND A4: IN VIVO CONSTRUCTION OF THE TETRASACCHARIDE BRANCHED CHAIN

With the kijanolide aglycone in hand, in vivo assays were attempted using *S. lividans*. A plasmid, pDIG, was constructed to generate TDP-L-digitoxose in *S. lividans* TK-24 strain. Each of the putative glycosyltransferase genes *kijA4*, *kijC1*, *kijC3, kijC4* and *kijD9* was cloned into the pWHM467-ermE* vector derivative and the genes were constitutively expressed in the digitoxose producing strain of *S. lividans*. Each of these *S. lividans* strains contained the pDIG plasmid, as well as one or more of the kijanimicin glycosyltransferase genes.

O-β-D-Kijanosyl-(1→17)-kijanolide (**13**) prepared as described in the previous section, was fed in a sucrose solution to multiple plates containing strains of *S. lividans* expressing the genes in the TDP-L-digitoxose biosynthetic pathway and one or more of the kijanimicin glycosyltransferases. [245] As a control, KijD9, the glycosyltransferase proposed to catalyze kijanose attachment was used, as it should not have activity towards the digitoxose substrate. Each strain was assayed for bioconversion of the kijanolide aglycone to the expected glycosylated products. Extraction and analysis of the glycosylated products by HPLC, MS and NMR demonstrated that glycosyltransferase KijC4 catalyzed the attachment of the first digitoxose to the aglycone (**13**) generating a new compound **14**, as shown in Figure 3.5.

Product **14** was isolated and subjected to MS and NMR analysis. These experiments were repeated and slightly different HPLC conditions were used to optimize the separation between compounds **13** and **14** (Figure 3.6).

3.5 KijC4 Catalyzes the First L-digitoxose Attachment to the Kijanolide Aglycon (**13**)

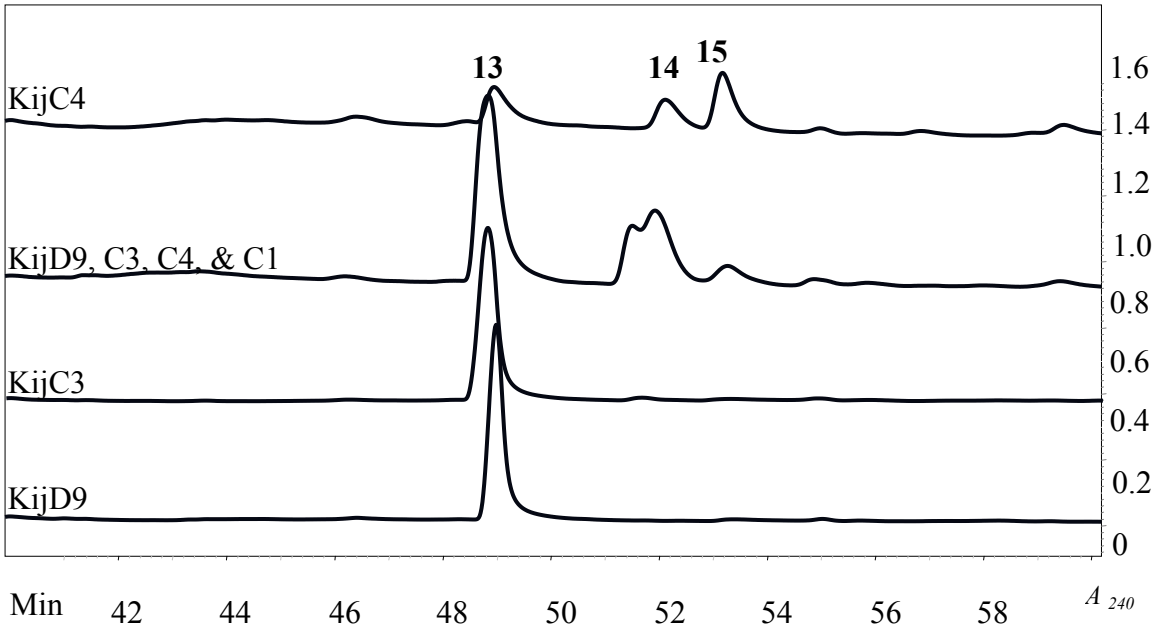


HPLC traces of *S. lividans* extracts showing the bioconversion of 13 to 14. Bacterial strains containing glycosyltransferases KijC4, KijC3, KijC1, KijA4 and KijD9 were fed 5 µg/mL agar (**13**) and grown for 72-128 h. Plates were extracted and kijanimicin compounds recovered by flash chromatography, concentrated and analyzed by HPLC.

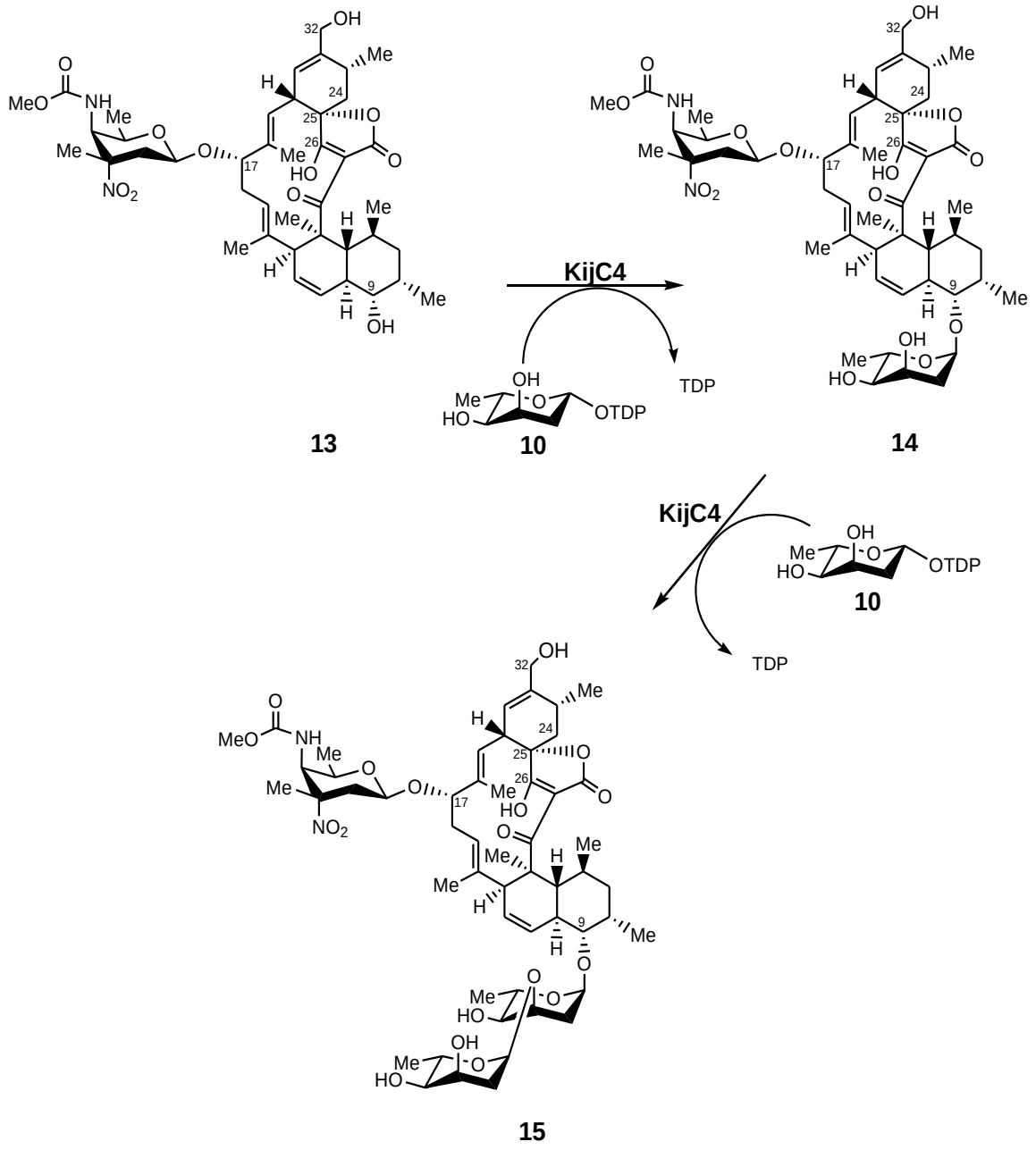
What was unexpected and surprising is that the protein KijC4 appears to also catalyze a second digitoxose addition, albeit at very low efficiency. These reactions were repeated several times, and the outcomes were reproducible. Interestingly, KijC4 not only transfers the first 2 digitoxose moieties to the kijanolide aglycone, by appears to be

capable of catalyzing the transfer of a third digitoxose. The one- (**14**) and two-sugar (**15**) containing products shown in figure 3.6 were analyzed by MS and ^{1}H-NMR.

3.6 KijC4 Catalyzes the Successive Transfer of Two TDP-L-digitoxose Moieties *in vivo*



HPLC traces of *S. lividans* extracts showing the bioconversion of 13 (aglycone) to 14 (one digitoxose) and 15 (two digitoxose). Bacterial strains containing GTs KijC4, C3, D9, and a construct expressing multiple GTs (D9, C4, C3 and C1) were fed 5 µg/mL agar (**13**) and grown for 72-128 h. Plates were extracted and kijanimicin compounds recovered by flash chromatography, concentrated and analyzed by HPLC.



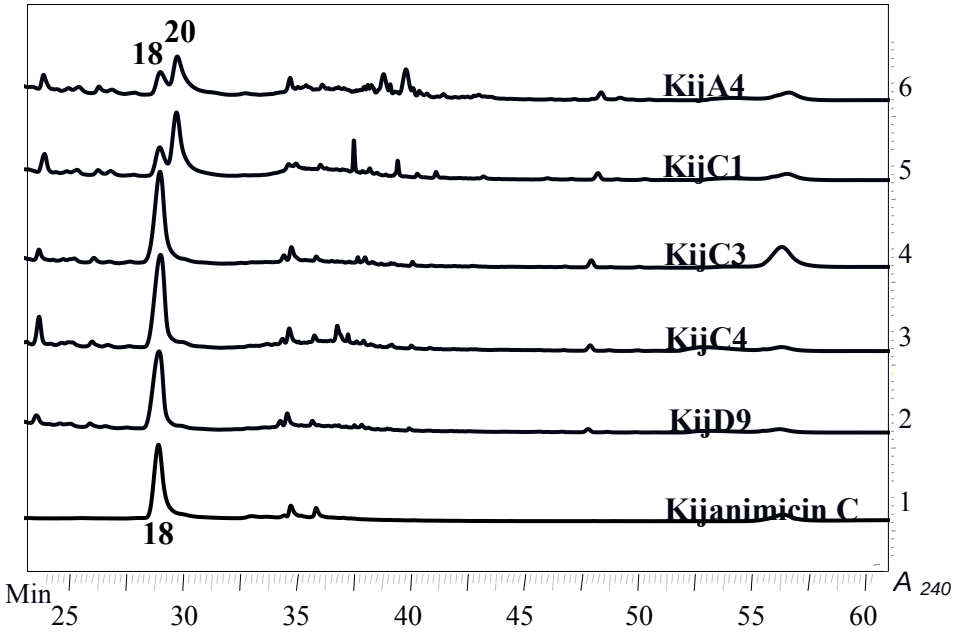
13 **10** **14** **10** **15**

The new compound (**14**) was found to have an (m/e), [M-H]; 911.458, calculated for $C_{48}H_{68}N_2O_{15}$: (m/e), [M-H] 911.454, while compound (**15**) was found to have an (m/e), [M-H]; 1041.518, calculated for $C_{54}H_{78}N_2O_{18}$: (m/e), [M-H] 1041.517. Further characterization by NMR confirmed the structures for the 1- and 2-sugar compounds as shown in Scheme 3.3. As mentioned previously, a 3-sugar compound (**16a** or **16b**) was isolated at very low yields. It was postulated that KijC4 is responsible for the attachment of the first two digitoxosyl moieties and also catalyzes the coupling of a third digitoxose in the absence of the real glycosyltransferase involved in the transfer of the third digitoxose moiety, which at this stage we reasoned could be either KijC3, KijC1 or KijA4.

Examining compound **15**, it was noted that this compound was very similar to kijanimicin C (**18**), differing only in the presence or absence of the hydroxyl group at the C-32 position. The feeding experiments were repeated in an attempt to identify the glycosyltransferases responsible for the subsequent digitoxose attachments. For these feeding experiments, kijanimicin C (**18**) was fed to plates containing strains of *S. lividans* containing pDIG and plasmids with each kijanimicin glycosyltransferase. We expected to observe the formation of **16a** or **16b** in one of these experiments with KijC1, KijC3 or KijA4, which would reveal the identity of the glycosyltransferase catalyzing the third digitoxose attachment.

When results from these experiments were analyzed, it was discovered that the putative glycosyltransferases KijC1 and KijA4 have apparent glycosidase activity towards kijanimicin C, and do not demonstrate glycosyltransferase activity, as shown in Figure 3.7. Note that a different HPLC program was utilized in this set of experiments to ensure separation between the expected 2-, 3-, and potential 4-sugar compounds.

3.7 KijC1 and KijA4 Demonstrate Glycosidase Activity Toward Kijanimicin C (**18**)

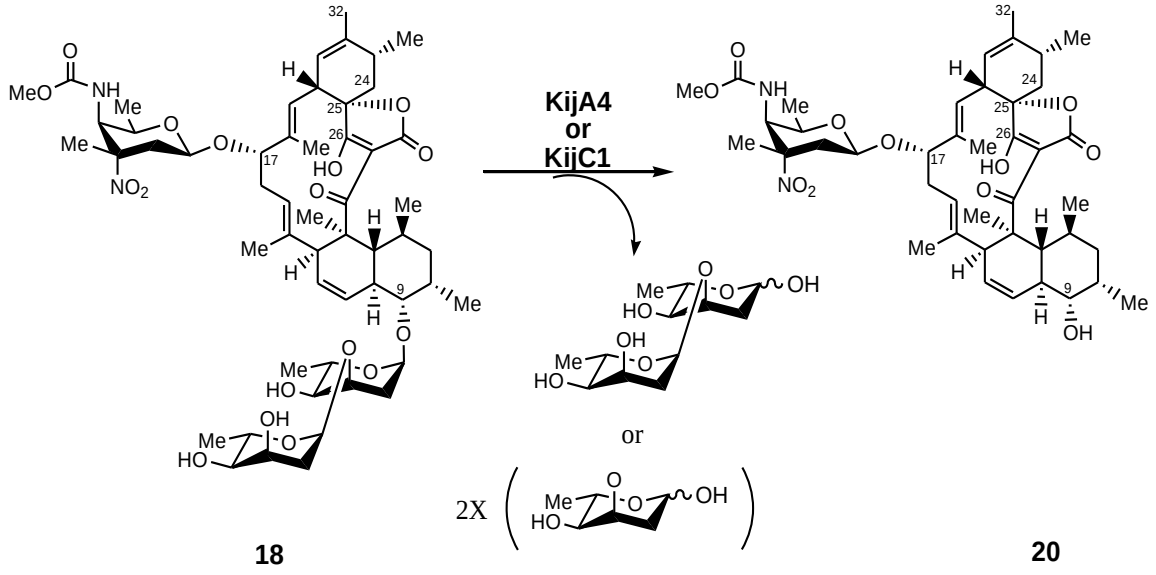


HPLC traces of *S. lividans* extracts showing the bioconversion of **18** (kijanimicin C) to **20** (32-deoxykijanolide). Bacterial strains containing glycosyltransferases KijC4, KijC3, KijC1, KijA4 and KijD9 were fed 5 µg/mL agar (**18**) and grown for 72-128 h. Plates were extracted and kijanimicin compounds recovered by flash chromatography, concentrated and analyzed by HPLC. Authentic standard for (**18**) shown in the bottom trace.

The substrate, kijanimicin C (**18**) with (m/e), [M-H] 1025.525 had previously been characterized by ^{1}H and ^{13}C-NMR. In strains producing the KijC1 and KijA4 proteins, a new peak (**20**) was formed. We postulated that the new peaks could be **16a** and/or **16b,** leaving KijC3 as a non-functional glycosyltransferase. Upon analysis by MS

and NMR, it was found that the new peak was not derived from the formation of 3-sugar compound (**16a** or **16b**) as we expected, but resulted from the formation of a new compound, *O*-β-D-kijanosyl-(1→17)-32-deoxykijanolide (**20**).

3.4 KijA4 and KijC1 Catalyzed Reactions

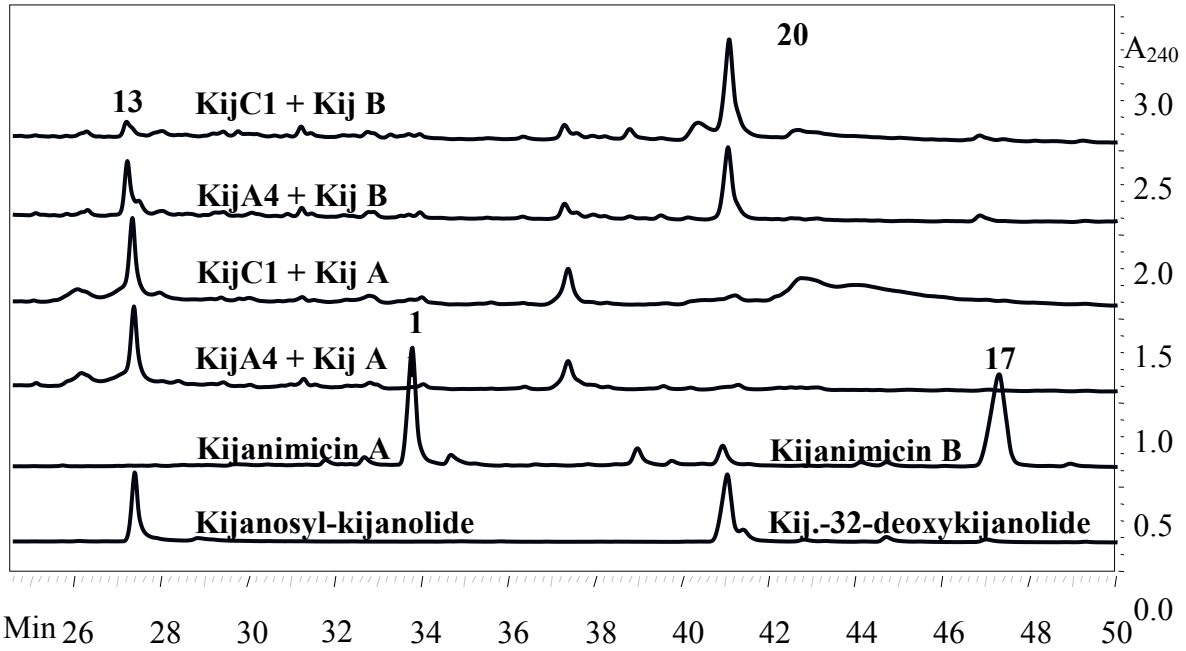


These results were quite disparate from what was expected. In the originally proposed pathway, each of 5 glycosyltransferases catalyzes the transfer of 1 of the 5 sugars to the kijanolide aglycone. Moreover, we reasoned that KijD9, due to its divergence from all other glycosyltransferases in the pathway is responsible for the coupling of the kijanose moiety to the kijanolide aglycone. The other four proteins (KijA4, KijC1, KijC3 and KijC4) are responsible for the transfer of each of the four

remaining digitoxose moieties to the kijanosyl-kijanolide (**13**). However, the above observation called for a revision of our previously proposed pathway.

To examine the glycosidase function of KijA4 and KijC1, strains containing genes for *kijC1* and *kijA4* expression were fed kijanimicin A (**1**) or kijanimicin B (**13**) and processed as previously described. If these proteins were truly glycosidases, KijA4 and KijC1 would remove digitoxose moieties from all kijanimicins A, B and C.

3.8 Glycosidase Activity of KijC1 and KijA4 Toward Kijanimicins A and B

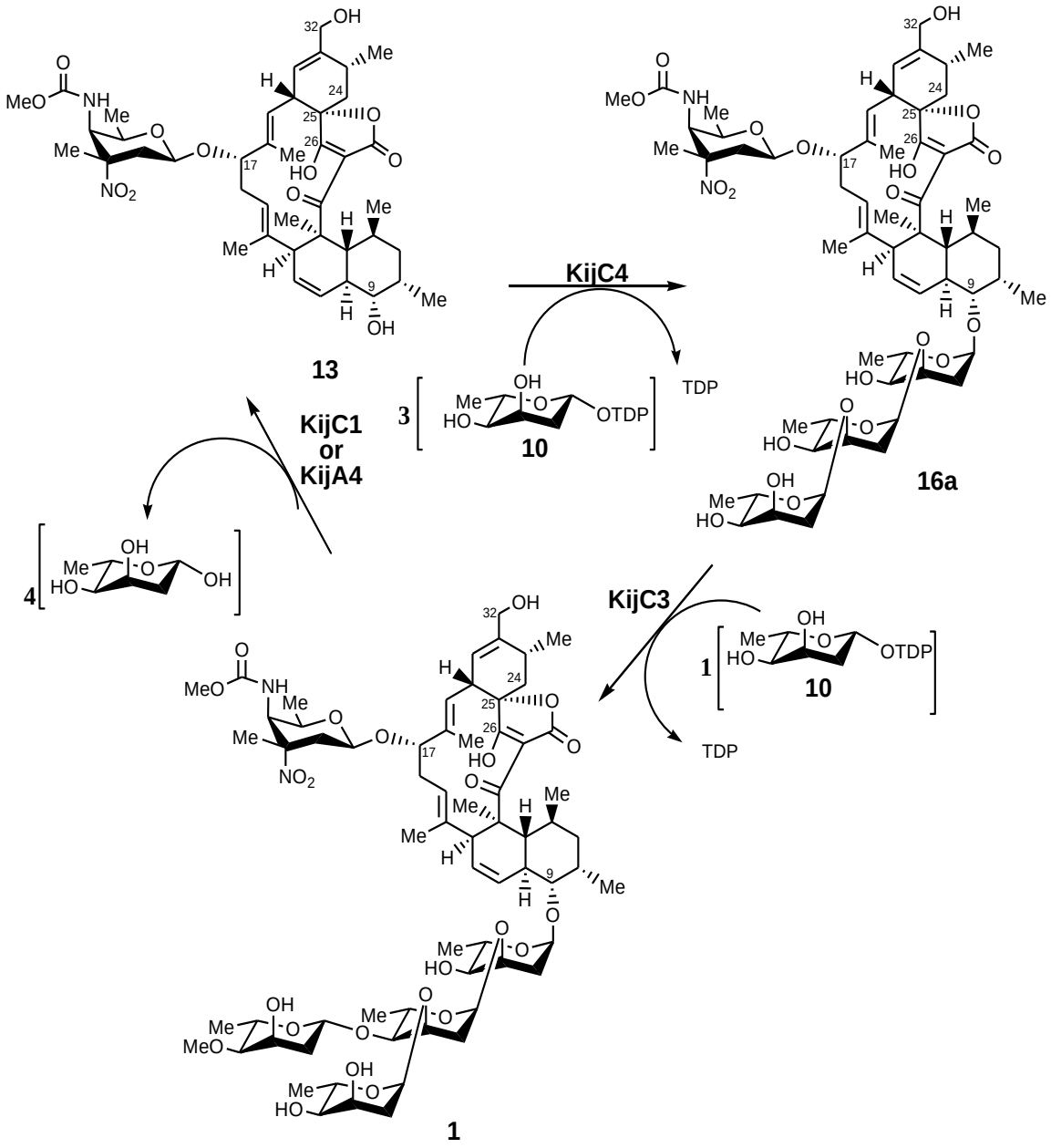


HPLC traces of *S. lividans* extracts showing the bioconversion of kijanimicin A (**1)** and kijanimicin B (**17)** to kijanosyl-kijanolide (**13**) and 32-deoxykijanolide (**20**), respectively. Bacterial strains containing putative GTs KijC1 and KijA4 were fed 5 µg/mL agar (**1** or **17**) and grown for 72-128 h. Plates were extracted and kijanimicin compounds recovered by flash chromatography, concentrated and analyzed by HPLC. Authentic standards for (**1**, **17**, **13** and **20**) are shown in the bottom two traces.

The feeding experiments clearly showed that KijC1 and KijA4 remove the branched chain 4-digitoxose structure from Kijanimicin A (**1**) and Kijanimicin B (**17**). These reactions yielded compounds *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**) and *O*-β-D-kijanosyl-(1→17)-32-deoxykijanolide (**20**), which were clearly visible on the HPLC traces, as seen in Figure 3.8. Thus, the results clearly demonstrate that KijA4 and KijC1 are in fact glycosidases and not glycosyltransferases as previously proposed. The results of the *in vivo* experiments with KijC4, KijC1 and KijA4 prompted a reevaluation of the proposed biosynthetic pathway, as shown in Scheme 3.5.

In the revised biosynthetic pathway, KijC4 transfers the first 3 L-digitoxose moieties to kijanolide in a successive fashion, and the transfer of remaining digitoxose moiety is catalyzed by KijC3. In this pathway, KijC4 functions as a processive glycosyltransferase, while KijC1 and KijA4 function as glycosidases. It should be noted that 3 of the digitoxosyl moieties have an α(1-3) linkage, while the branching digitoxose has a β(1-4) linkage. This fact is consistent with the proposed role of KijC4 which catalyzes the first 3 sugar additions, (for an unknown reason the 3-sugar intermediate evaded isolation). Logically, the 3 α(1-3) linkages could be catalyzed by a single processive enzyme, while the remaining β(1-4) linkage formation is catalyzed by a different glycosyltransferase. The formation of both α and β glycosidic bonds mediated by a single enzyme would require that enzyme to have the ability to catalyze both an inverting and a retaining reaction in one active site. There is no precedent for a glycosyltransferase enzyme with such ability to catalyze both S_N1 and S_N2 type reactions in one active site. Thus, it is much more likely that one enzyme (KijC4) catalyzes the α (inverting) glycosyltransfer reactions in a processive manner while a different enzyme (KijC3) catalyzes the β (retaining) glycosyltransfer.

3.5 Revised Kijanimicin Glycosylation Proposed Pathway



These two proteins have 53% pairwise identity, yet must catalyze completely different reactions, an inverting S_N2 and retaining S_N1. It is intriguing to discover

glycosyltransferases in a single gene cluster sharing such high sequence identity while employing completely different mechanisms.

Glycosyltransferases responsible for multiple additions are known. For example, heterologous expression of the landomycin cluster in a *S. fradiae* mutant led to the production of landomycin A demonstrating that four LanGTs are responsible for the formation and attachment of six sugars.[307] However, the glycosyltransferases LanGT4 and LanGT1 that are responsible for the attachment of two sugars each, do not act in a processive fashion.[71] In this case, LanGT1 is responsible for the addition of the second and fifth sugars, while LanGT4 catalyzes the addition of the third and sixth sugars. The intermediate fourth sugar is added by LanGT3, while the initial glycosyltransfer is catalyzed by LanGT2.[71] Thus, although these glycosyltransferases are known to catalyze the addition of multiple sugars, they do not catalyze multiple sequential additions in a processive fashion, as KijC4 appears to do.

Processive glycosyltransferases such as PglH in *C. jejuni* have a single active site and catalyze successive reactions by slipping the product through its active site. [308] This glycosyltransferase is interesting to examine as kinetic studies by the Imperiali group have suggested a counting mechanism by which the protein adds three UDP-GalNAc residues to the growing glycan chain to complete the final product. [308] This model may provide insights into the mechanism of the KijC4 glycosyltransferase. PglH has increased binding affinity as the length of the glycan chain increases, which in turn slows the rate of the glycosyltransferase reaction. [308] Thus, the initial glycosyltransfer reactions are quite slow due to low binding affinity for the substrate, and the latter reactions are slow due to the low catalytic rate of the enzyme. [308] Under high substrate (aglycone) conditions, the 1-sugar intermediate products dissociated at a high rate from the enzyme,

while 2- and 3-sugar compounds were retained longer by PglH. [308] In this glycosylation pathway, the product is removed from the cell thus preventing product inhibition. [308]

Again, it is interesting to note that the 3-sugar compound could not be found in the feeding experiments or the extract from the original *A. kijaniata* culture, although in the PglH system, all intermediates were observed *in vitro*. Perhaps a high affinity for KijC4 towards the 3-sugar compound can explain this apparent absence in the experimental results. In *A. kijaniata* cultures, the 3-sugar compound (**16a**) was not observed, perhaps due to the presence of KijC3 in the system. If KijC3 is a fast enzyme, there would be no detectable accumulation of its substrate (**16a**). In some cases, we were able to recover a compound with the same (m/e) as the predicted 3-sugar compound, but were never able to characterize it by NMR.

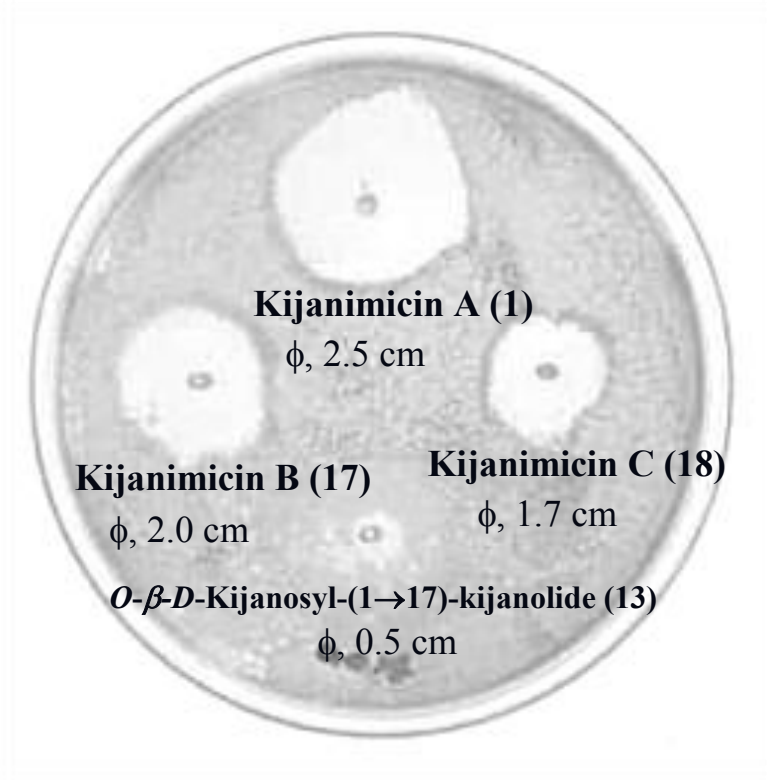
With the new proposal, feeding experiments involving KijC3 were undertaken, in an attempt to generate the full-length product kijanimicin A (**1**) from kijanimicin C (**18**) and the two sugar compound **15**. If KijC3 is responsible for the addition of the latter 2 digitoxosyl moieties, kijanimicin A (**1**) should be produced. If KijC3 is responsible for only the β(1-4) addition, the starting material should be recovered, and no new products seen in the HPLC traces. Briefly, compound **15** (2SC) or kijanimicin C (**18**) was fed to strains of *S. lividans* containing the pDIG and *kijC3* expression plasmids. The results were as we expected, no new compound was added, thus encouraging the belief that the substrate of KijC3 is a 3-sugar intermediate. No new compounds were observed in the HPLC traces or by TLC analysis of the crude extracts. In some cases, compound **15** or kijanimicin C (**18**) could not be recovered from the plate extracts and disappeared completely from HPLC traces. In other cases, the original compounds were recovered, and as expected, additional glycosylation was not observed.

3.3.4 S. LIVIDANS SUSCEPTIBILITY TO KIJANIMICINS

The inability to achieve more than 2 digitoxose additions using the heterologous expression system in *S. lividans* prevented further *in vivo* work in this platform. We reasoned that perhaps the failure to detect the formation of the 3-sugar or 4-sugar (kij A) compound could be due to toxicity. If kijanimicin A (**1**) is toxic to *S. lividans*, then the generation of the compound would cause cell death and thus, no significant amounts of product would ever be accumulated in the feeding experiment extractions. Thus, the susceptibility of the host bacteria to kijanimicin compounds was tested. By growing *S. lividans* on an agar plate and applying 40 μg of each compound (kijanimicin A (**1**), B (**17**), C (**18**) and kij-kij (**13**)) to the plate, significant zones of growth inhibition were observed, as seen in Figure 3.9. Clearly, the more digitoxosyl moieties are added to the kijanolide aglycone, the more toxic the compound becomes. Thus, *in vivo* experiments could not be continued successfully in this host, because the generation of compound **15** (2-SC) is detrimental to the host.

This may also explain why the 3-sugar intermediate had not been observed in *S. lividans* feeding experiments. Perhaps strains producing significant quantities of the compound were killed due to toxicity. Thus, it is likely that KijC4 does in fact attach the first 3 digitoxosyl moieties in an α(1-3) linkage, but due to toxicity, only the products carrying the first 2 sugars can be observed. Obviously, without the 3-sugar compound in hand, we cannot examine its toxicity, although the effect can be inferred from the results of kijanimicins B (**17**) and C (**18**), which are 4- and 2-sugar compounds respectively, that the 3-sugar compound could indeed possess an inherent level of toxicity towards the heterologous host. However, the lack of accumulation of the 3-sugar compound (**16a**) may also be attributed to product inhibition or high binding affinity towards KijC4.

3.9 *S. lividans* Susceptibility to Kijanimicins A, B, C and Kijanosyl-Kijanolide.



S. lividans growth inhibition due to kijanimicin compounds. Mycelia was plated on agar plates and compounds were applied in a radius of 0.1 cm. After incubation at 30 ºC, zones of growth inhibition were measured. Growth inhibition was found to correlate with the presence of the C-32 hydroxy group and the number of L-digitoxose moieties present.

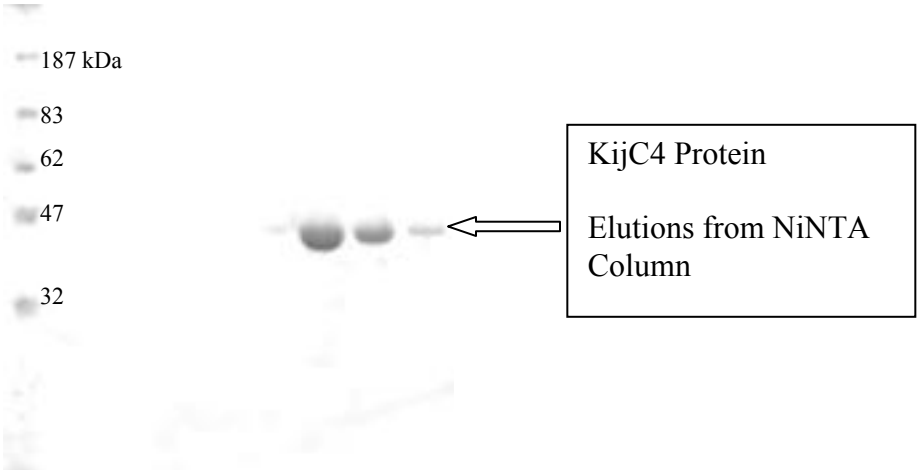
3.4 KIJANIMICIN GLYCOSYLTRANSFERASES C1, C3, C4 AND A4: IN VITRO CONSTRUCTION OF THE TETRASACCHARIDE BRANCHED CHAIN

To continue these experiments, a shift to an *in vitro* platform was attempted due to the toxicity of kijanimicins towards the host *S. lividans*. A concern regarding *in vitro* work was the instability of TDP-L-digitoxose previously observed in the biosynthetic studies carried out in Chapter 2.

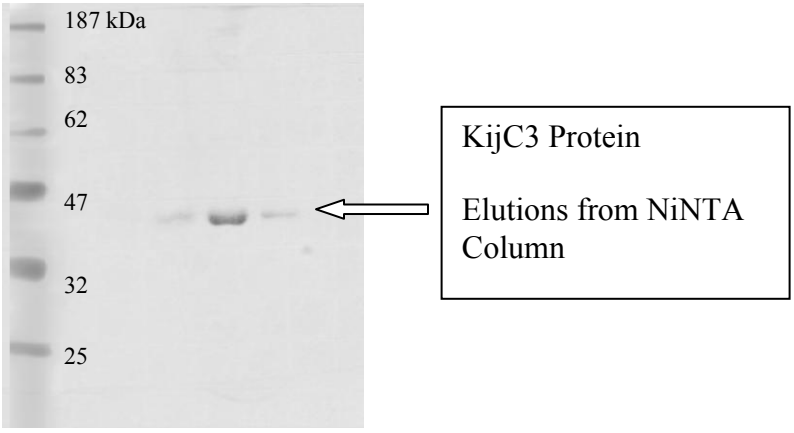
To prepare the kijanimicin glycosyltransferases, each of the putative biosynthetic genes (*kijA4*, *kijC1*, *kijC3*, *kijC4* and *kijD9)* was amplified by PCR and cloned into the pET28b vector to generate N-terminal His_6-tagged proteins. Glycosyltransferase genes *kijD9*, *kijA4*, *kijC1*, *kijC3* and *kijC4* were also introduced into a pWHM467-ermE* vector and overexpressed in *S. lividans*. Glycosyltransferase genes *kijA4* and *kijC1* were subsequently overexpressed in *E. coli* using the pET28b constructs. Culture conditions were optimized for each enzyme to obtain maximum amounts of soluble protein. The gene products were purified by Ni-NTA affinity chromatography, to near homogeneity. Purification of these proteins from *S. lividans* afforded very clean proteins, with little to no contaminants, as shown in Figures 3.10-3.14. KijA4 and KijC1 from *E. coli* preparations were not as pure. When the remaining glycosyltransferases (KijC4, KijC3, and KijD9) were expressed in *E. coli*, expression and solubility problems were encountered.

For these assays, several substrates were employed. Kijanimicins A (**1**), B (**17**) and C (**18**) were obtained through organic extractions of *A. kijaniata* culture media and subsequent purification as described in sections 3.2.9 and 3.2.10.[6] Compound **15** was obtained following the procedures in section 3.2.15. To obtain the kijanolide substrate **13**, the L-digitoxose tetrasaccharide chain was removed by mild acid hydrolysis of kijanimicin A.[245]

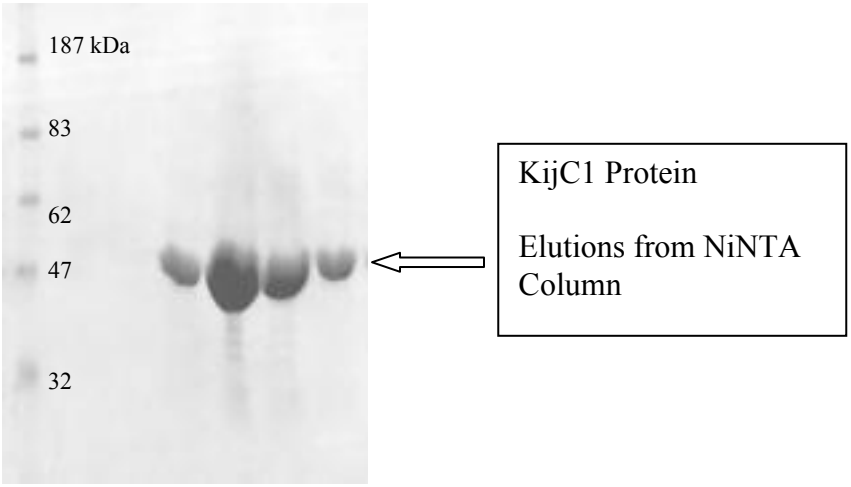
3.10 KijC4 Protein Purification from *S. lividans*



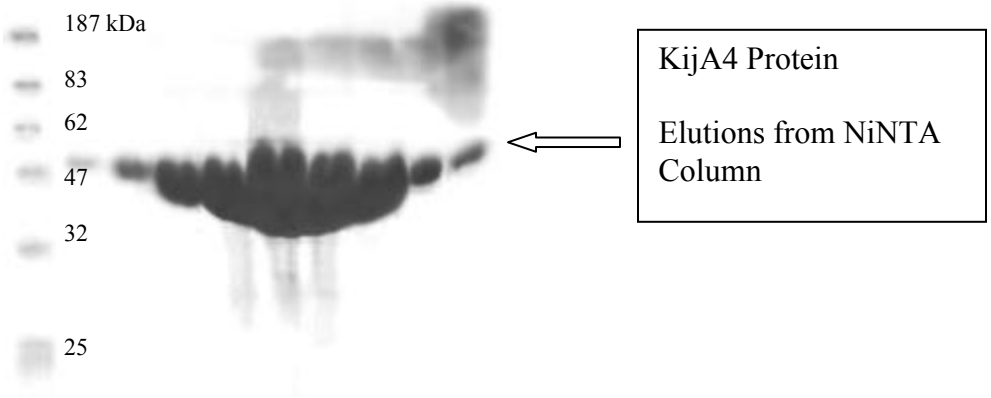
3.11 KijC3 Protein Purification from *S. lividans*



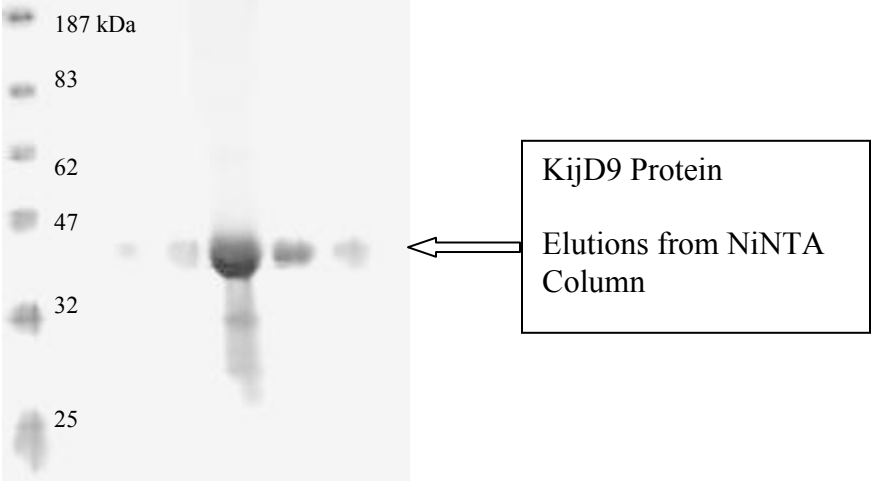
3.12 KijC1 Protein Purification from *S. lividans*



3.13 KijA4 Protein Purification from *S. lividans*



3.14 KijD9 Protein Purification from *S. lividans*

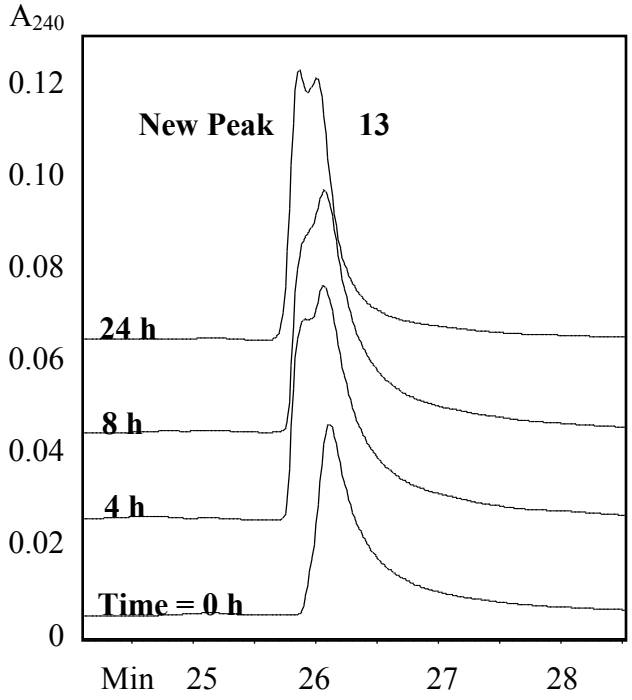


Many experiments were performed to elucidate the function of these glycosyltransferases. Briefly, each glycosyltransferase protein was incubated with crude TDP-L-digitoxose and kijanolide substrate **13** to generate the product *in vitro.* After room temperature incubation, the proteins were removed and the samples were analyzed with an HPLC equipped with a C-18 column monitoring at 240 nm, using the previously described elution program. The results were disappointing, and no new product was seen in these assays. Assay conditions were altered, using different pH and protein concentrations, the experiments were repeated to no avail. Due to the instability of the digitoxose substrate, it is possible that these reactions were doomed to failure as the digitoxose degrades rapidly at room temperature.

To overcome the inherent instability of the TDP-L-digitoxose sugar, assays were designed such that the sugar would be biosynthesized *in situ*. Glycosyltransferases were incubated with kijanolide (**13**), (**14**) or (**15**) substrates in reactions with TDP-4-keto-2,6-

dideoxy-D-glucose, KijD11 and KijC2. These reactions were monitored by TLC and analytical scale HPLC, employing the conditions described above.

3.15 Kijanimicin Glycosyltransferase C4 Activity Towards *O*-β-D-kijanosyl-(1→17)-kijanolide (**13**)



GTC4 *in vitro* reactions demonstrating the conversion of **13** to a new peak. Reactions consisted of the following: 1 mM TDP-4-keto-2,6-dideoxy-D-glucose, 10 nM KijD11, 10μM KijC2, 30 μM KijC4 10 mM $MgCl_2$, 1.2 mM kij-kij (**13**) (in MeOH), 10% glycerol, 1.4 mM NADH and 50 mM Tris (pH 9.0) in a final reaction volume of 200 μL. Aliquots (40 μL) were removed from the reaction at various time points (0 h, 2 h, 4 h, 8h, overnight) and flash frozen. HPLC analysis employed a C-18 column, using the eluents and gradient program previously described.

Of all the proteins tested, only KijC4 demonstrated activity towards the kijanolide substrate (**13**), as shown in Figure 3.15. These reactions were repeated and the results

confirmed, a new peak appears and increases over time when (**13**) is incubated in the presence of KijC4, but not in the presence of other glycosyltransferases (KijD9, KijC3, KijC1, KijA4), which supports the previous *in vivo* results suggesting that KijC4 is in fact responsible for the first glycosylation steps, however we were unable to obtain MS data for the expected product. This reaction was repeated, in an attempt to observe two digitoxose additions. We attempted multiple conditions, and for the aglycone used the previously characterized compounds **14** and **15** as substrates, as well as freshly generated **14**. Attachment of the second digitoxose was not observed under any of the conditions that we attempted. Reasoning that the instability of the TDP-L-digitoxose (**10**) substrate was a limiting factor in these assays, another method for obtaining the sugar was undertaken.

3.4.2 REVERSE GLYCOSYLTRANSFERASE REACTIONS

The Thorson group, among others, has discovered that in the presence of excess NDP, some glycosyltransferases will catalyze a reverse reaction and transfer the sugar from glycosylated product to NDP, thus creating the NDP-activated sugar and aglycone.[299, 309, 310] Our reactions were designed following the method of Thorson, in which an excess of TDP is incubated with glycosyltransferase and glycosylated product.[299, 309, 310] Under these conditions we hoped that the kijanimicin glycosyltransferases would catalyze the reverse reaction, transferring the digitoxose sugar from the glycosylated product kijanimicin A (**1**) to TDP, generating a TDP-L-digitoxose (**10**) and an aglycone with one (or more) fewer appended sugar. Thorson and co-workers also demonstrated that reverse reactions including a second glycosyltransferase with substrate specificity for the NDP-activated sugar and target aglycone can be used to

generate a new glycosylated natural product using the NDP-activated sugar generated *in situ*. [299, 309, 310]

Previous feeding experiments indicated that KijA4 and KijC1 were not glycosyltransferases, but instead are novel glycosidases. *In vivo*, these proteins remove the digitoxosyl branched sugar chain from kijanimicins A, B and C. The reversibility of glycosidases has never been demonstrated. We were interested to know if these proteins were capable of catalyzing reverse reactions to generate TDP-L-digitoxose (**10**). Due to the difficulty we encountered with the production and purification of the TDP-L-digitoxose, and its inherent instability, the ability to generate the TDP-activated sugar *in situ* would allow us to conduct extensive glycosyltransferase assays *in vitro*. We also examined the ability of KijC4 to catalyze the reverse glycosyltransferase reactions.

KijC1 and KijA4 from *E. coli* and KijC1, KijA4 and KijC4 from *S. lividans* preparations were incubated with kijanimicin A overnight in the presence of TDP to determine if they could catalyze the reverse reaction as seen *in vivo*. After incubation, the samples were divided and analyzed using the Carbopac PA-1 and C-18 HPLC columns to assess both the formation of TDP-L-digitoxose (**10**), and the reduction of kijanimicin A (**1**), respectively. In these experiments, the reverse reactions were not observed, nor did we observe a decrease in the amount of kijanimicin A (**1**). These results indicated that most likely the KijC1 and A4 proteins are not active *in vitro*, while KijC4 does not appear to catalyze the reverse reaction with the kijanimicin A substrate. It is may be that the addition of the β-linked sugar in kijanimicin A precludes the possibility of the reverse reaction of KijC4. In terms of the failure of KijA4 and KijC1 to catalyze the reverse reaction, it is unclear whether the purification conditions or reaction conditions resulted in protein inactivity, although the possibility exists that the proteins were folded incorrectly during heterologous expression. There is no way to examine folded state of

the proteins and compare it to the native proteins, as we were unable to manipulate the DNA of *A. kijaniata*. Thus, isolation of the natively folded protein is not possible at this stage.

These results call into question the glycosidase activity observed in the *S. lividans* feeding experiments. If the glycosyltransferase proteins are folded incorrectly in the heterologous host, perhaps the glycosidase activity is merely the result of misfolding. It is tempting to retain the original biosynthetic proposal that each glycosyltransferase (C4, C3, C1, and A4) is responsible for the attachment of only one digitoxosyl moiety. However, the processive activity of glycosyltransferase KijC4 must be accounted for in some fashion. The *in vivo* results clearly indicate that KijC4 is a processive glycosyltransferase responsible for the attachment of the first 2 digitoxosyl moieties in an α(1-3) linkage.

What, then is the function of having two glycosyltransferases that function as glycosidases? It is possible that these glycosidases are involved in self-resistance to kijanimicins. As the bioassay shown in Figure 3.9 clearly demonstrates, toxicity increases with the addition of more digitoxosyl moieties to the tetrasaccharide branched chain. Thus, an obvious mechanism of self-resistance employed by *A. kijaniata* to avoid this toxicity would be the removal of the digitoxosyl moieties from the kijanimicin biosynthetic products. Glycosylation or glycosidation as a self-resistance mechanism is not common, but it is not unknown, either. One clear example of a resistance mechanism based on glycosylation is in the pikromycin/methymycin/neomethymycin biosynthetic pathways of *S. venezuelae*, and involves DesG.[311] This glycosyltransferase attaches a glucose moiety to the natural products.[311] DesG thus "glucosylates" the natural products in the cell, protecting the host from an excess of the natural products in the cell.[311] Similar phenomena has also been observed in the oleandomycin producing strains in

which OleD catalyzes the glucosylation of oleandomycin as part of a self-resistance mechanism.[312]

3.5 CONCLUSIONS

Herein, we report the results of our investigations of the glycosyltransferase genes of the kijanimicin biosynthetic pathway, and revisit the production of TDP-L-digitoxose both *in vitro* and *in vivo*. The TDP-L-digitoxose biosynthetic pathway was reconstructed *in vivo*. A plasmid (pDIG) containing all genes necessary for TDP-L-digitoxose biosynthesis was constructed and introduced to *S. lividans*. Each putative glycosyltransferase gene was co-expressed in the stain containing the pDIG plasmid. These strains were fed various kijanolides and the results show that KijC4 is clearly responsible for the first two digitoxosyl α(1-3) attachments. Furthermore it is also likely responsible for the third digitoxosyl α(1-3) attachment in a processive manner. The successful addition of digitoxosyl moieties to appropriate aglycones demonstrated *in vivo* further confirms that these deoxysugar genes identified and characterized in Chapter 2 are indeed responsible for the biosynthesis of TDP-L-digitoxose. The putative glycosyltransferases KijC1 and KijA4 were found to have glycosidase activity *in vivo* towards kijanimicins A, B and C, leading to the proposal of their involvement in self-resistance. The KijC3 glycosyltransferase is hypothesized to catalyze the remaining β(1-4) linkage. Unfortunately, this hypothesis cannot be examined at the present time due to a lack of the proposed substrate. Finally, we reevaluated the proposed glycosylation scheme for kijanimicin and developed a new biosynthetic proposal for kijanimicin glycosylation as shown in Figure 3.5. If the 3-sugar compound becomes available in the future, the function of KijC3 needs be examined. Investigation of the glycosyltransferase

ability to accept a linear aglycone substrate can also be examined if the linear aglycone is synthesized.

4 Eukaryotic Glycosylation Studies: *Identification of H. sapiens Candidates for the C-mannosyltransferase and their expression in E. coli and S. cerevisiae*

4.1 INTRODUCTION

C-mannosylation is a novel protein posttranslational modification discovered on RNase 2 in 1994 by Hofsteenge et al.[229] Subsequent studies revealed that the recognition sequence for this type of protein modification is a W-X-X-W motif, with *C*-mannosylation occurring on the first tryptophan residue.[313] Replacement of the second tryptophan by other amino acids eliminated or substantially reduced the level of *C*-mannosylation.[314] Accumulated evidence showed that this modification is enzyme catalyzed and occurs in *Caenorhabditis elegans*, amphibians, birds and mammals, but not *Escherichia coli*, insects, or yeast.[314, 315] The presence of *C*-mannosylation in fungal species other than *Saccharomyces* has not been investigated. More recent studies have found that marine organisms such as mollusks produce *C*-mannosylated natural products, although it is not clear if these small molecules are produced by the mollusks themselves, or by marine microorganisms that may colonize them.[316]

In vitro assays demonstrated that this modification occurs in the lumen of the endoplasmic reticulum (ER) leading to the hypothesis that many *C*-mannosylated proteins are secretory or transmembrane in nature. [315] Over twelve thousand proteins in the mammalian genome contain this W-X-X-W recognition sequences. Limiting this set by cellular localization, 336 of these proteins have been proposed to be potentially *C*-mannosylated.[314] These proteins have diverse functions, from channels, receptors and transporters to immunoglobulins and enzymes. Many more proteins that carry this motif

are of "unknown function" or are "hypothetical" proteins. More recent research in this area focused on immune system proteins from the complement pathway, which have been found to be *C*-mannosylated.[317] Thus far, 24 modification sites have been experimentally determined on seven proteins including properdin, IL-12b, RNase2, C6, C7, C8a/C8b, and C9, as shown in Table 4.1. [317,318]

4.1 Known *C*-Mannosylation Modifications on Human Proteins

Gene Name	Position	Signal Subsequence	Protein Description
C7	506-509	GGWSCWSSWSPCVQG	Complement component C7 precursor
C8A	44-47	CQLSNWSEWTDCFPC	Complement component C8 alpha chain precursor (Complement component 8 α subunit)
C8A	542-545	KADGSWSCWSSWSVC	Complement component C8 alpha chain precursor (Complement component 8 α subunit)
C8A	545-548	GSWSCWSSWSVCRAG	Complement component C8 alpha chain precursor (Complement component 8 α subunit)
C8B	70-73	CELSSWSSWTTCDPC	Complement component C8 beta chain precursor (Complement component 8 β subunit)
C8B	548-551	PIDGKWNCWSNWSSC	Complement component C8 beta chain precursor (Complement component 8 β subunit)
C8B	551-554	GKWNCWSNWSSCSGR	Complement component C8 beta chain precursor (Complement component 8 β subunit)
C9	48-51	CRMSPWSEWSQCDPC	Complement component C9 precursor
RNASE2 (EDN, RNS2)	34-37	PPQFTWAQWFETQHI	Nonsecretory ribonuclease precursor (EC 3.1.27.5) (Ribonuclease US) (Eosinophil-derived neurotoxin) (RNase UpI-2) (Ribonuclease 2) (RNase 2)

IL12B (NKSF2)	319-322	YYSSSWSEWASVPCS	Interleukin-12 beta chain precursor (IL-12B) (Cytotoxic lymphocyte maturation factor 40 kDa subunit) (CLMF p40) (NK cell stimulatory factor chain 2) (NKSF2).
PFC	80-83	CRSPRWSLWSTWAPC	Properdin precursor (Factor P)
PFC	83-86	PRWSLWSTWAPCSVT	Properdin precursor (Factor P)
PFC	139-142	PEMGGWSGWGPWEPC	Properdin precursor (Factor P)
PFC	142-145	GGWSGWGPWEPCSVT	Properdin precursor (Factor P)
PFC	196-199	PTHGAWATWGPWTPC	Properdin precursor (Factor P)
PFC	199-202	GAWATWGPWTPCSAS	Properdin precursor (Factor P)
PFC	260-263	PVAGGWGPWGPVSPC	Properdin precursor (Factor P)
PFC	318-321	PVDGEWDSWGEWSPC	Properdin precursor (Factor P)
PFC	321-324	GEWDSWGEWSPCIRR	Properdin precursor (Factor P)
PFC	382-385	PLKGSWSEWSTWGLC	Properdin precursor (Factor P)
PFC	385-388	GSWSEWSTWGLCMPP	Properdin precursor (Factor P)
C6	29-32	CDHYAWTQWTSCSKT	Complement component C6 precursor
C6	568-571	AVDGQWGCWSSWSTC	Complement component C6 precursor
C6	571-574	GQWGCWSSWSTCDAT	Complement component C6 precursor

Cytokine receptors may also be *C*-mannosylated. MBL (mannose binding lectins), proteins which activate complement and bind to the outer membranes of some pathogens, have a potential *C*-mannosylation site, thus implicating the importance of this protein modification for the entire immune pathway. However MBL-C, a major mannose-binding lectin purified from mouse serum, does not bind *N*-biotinylated *C*-Man-Trp, based on an ELISA assay.[235] Concanavalin A, a lectin that binds p-*O*-mannose residues, also failed to bind *C*-Man-Trp.[235] These results imply that *C*-Man-Trp is not recognized by canonical lectins that bind to *O*-linked mannose, but may be recognized by other specific proteins associated with unknown biological functions.

The biological role of *C*-mannosylation has not been experimentally determined. The proposed roles of *C*-mannosylation include protein stability, control of protein

secretion, stabilization or orientation of tryptophan residues, and protein interaction in the extracellular matrix. For example, in the case of RNase2, mannose appears to interact with loop residues that stabilize the N-terminus of the protein and orient the indole ring of the modified tryptophan towards the RNase2 active site. For example, protein modeling of RNase2 suggests that the mannose moiety likely occupies a cavity adjacent to an *N*-terminal loop that is located just outside the active site of this protein. [232, 319] It has also been suggested that the hydrophilic nature of *C*-mannosylation orients the *C*-mannosylated tryptophan toward the surface of the modified protein, which may play structural or functional roles.[227] For example, properdin serves the function of a protein scaffold, stabilizing the C3 convertase (C3bBb) and protecting the protein from rapid inactivation.[317] In this case, the *C*-mannosylated residue is located on the surface of properdin. *C*-Mannosylation is not necessarily required for enzymatic activity (as in the case of RNase2), but mutation of either tryptophan in the *C*-mannosylation signal motif in the IL-2 receptor abolishes ligand binding.[227] It has previously been demonstrated that mutations of the ^{232}WSAWS^{236} motif in the erythropoietin receptor (EPOR) can result in strong inhibition of receptor surface expression due to defective intracellular transport.[320] However, *C*-mannosylation did not assist in the expression of EPOR on the cell surface or intracellular transport.[321] Clearly, much more work remains to be done to understand the potential roles of this novel modification and the cellular processes in which it may be involved.

The discovery of protein *C*-mannosylation is a significant contribution to the protein glycosylation field, as post-translational modifications are usually essential for the function of a given protein.[227] The ability to reproduce the correct glycosylation pattern is important for studying the functions of recombinant glycoproteins, which constitute over one-third of all approved therapeutic proteins.[322] To achieve the necessary

human-like glycosylation level, these proteins are generally produced using mammalian cell lines.[322] However, isolation of the desired glycoproteins from mammalian cell culture is lengthy, cumbersome and it is difficult to ensure homogeneity of the final product glycoforms. Thus, reconstructing human glycosylation pathways in cells that allow more control over the purity of the final product are highly desirable, since the extent and type of glycosylation have been found to have a marked impact on the therapeutic properties of many commercial therapeutic proteins.[322] It has been shown that hyperglycosylation of therapeutic proteins would increase long-term stability of these proteins *in vivo*, by increasing their serum half-life and cellular uptake. Cerezyme®, which is used in enzyme replacement therapy for the treatment of Gaucher's disease, is prepared recombinantly based on the human gene sequence. The requisite N-glycans are modified to expose α-mannosyl residues for enhanced liver (kupffer cell) and macrophage uptake.[323, 324] Natively glycosylated protein has a clearance rate of 21 min, while modified (terminal α-mannosyl) has a clearance rate of 3.2 min, demonstrating increased cellular uptake.[324] If it is true that many proteins in human serum are *C*-mannosylated, it is possible that therapeutic proteins produced as recombinants could be more efficacious if they were also *C*-mannosylated. Even if the target protein lacks native *C*-mannosylation, the addition of one or more mannose residues could increase therapeutic efficacy, as seen with erythropoietin (EPO).[325] *C*-mannosylation levels also appear to increase in specific tissues or cell types under hyperglycemic conditions, suggesting a pathological role for *C*-mannosylation abundance in the development of diabetic complications.[326]

Clearly the questions that remain to be answered outnumber known facts regarding the functions and roles of *C*-mannosylation. It has been reported that insect cells do not exhibit *C*-mannosylation, however, a secondary modification sequence,

termed a thrombospondin repeat sequence (WXXWXXWXXC), has also been identified, and is apparently recognized and *C*-mannosylated by both *Spodoptera frugiperda* Sf-9 and *Trichoplusia ni* Hi-5 insect cell lines, contradicting earlier reports.[228, 317] A bovine lens protein was also found to be *C*-mannosylated at an amino acid sequence that does not correspond to either the canonical or thrombospondin repeat sequences, leading to the proposal that there exist multiple classes of *C*-mannosyltransferase enzymes.[327] The first step in answering these myriad questions is to identify the enzyme(s) responsible for this unique glycosylation event.

In this chapter, we attempted to identify the putative *C*-mannosyltransferases by pull-down assays. We also attempted to identify the putative *C*-mannosyltransferases by analysis *in silico*, using protein *O*-mannosyltransferase as a BLAST query to isolate potential enzyme candidates. These candidates were cloned and we attempted to express the genes in *E. coli* and *S. cerevisiae* to test their enzymatic functions.

4.2 EXPERIMENTAL DETAILS

4.2.1 MATERIALS

4.2.1.1 Bacterial and Yeast Strains

E. coli strain DH5α was obtained from Bethesda Research Laboratories (Gaithersburg, MD). The overexpression hosts *E. coli* BL21, BL21(DE3), BL21(DE3)PLysS and BL21 RosettaII(DE3) were purchased from Novagen (Madison, WI). *Saccharomyces cerevisiae* yeast strain INVSc1 was acquired from Invitrogen (Carlsbad, CA).

4.2.1.2 Biochemicals

Enzymes and molecular weight standards used for the molecular cloning experiments were products of Invitrogen (Carlsbad, CA) or New England Biolabs (Beverly, MA). Agarose for DNA electrophoresis was purchased from Fisher Scientific (Pittsburgh, PA). Restriction digestion enzymes, calf intestinal alkaline phosphatase (CIAP), 100X bovine serum albumin (BSA), T4 ligase, and their respective buffers were products of New England Biolabs (Beverly, MA). Ni-NTA agarose and kits for DNA gel extraction and spin miniprep were obtained from Qiagen (Valencia, CA). The growth media components for bacteria and yeast were acquired from Becton Dickinson (Sparks, MD). Complete supplemental media lacking uracil and histidine (csm -ura-his) was a product of Bio-101 Systems/MP Biomedical, (Solon, OH). Antibiotics, amino acids, nucleosides and chemicals such as isopropyl-β-D-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were products of Sigma-Aldrich Chemicals, Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). All reagents for SDS-PAGE were purchased from Bio-Rad (Hercules, CA), with the exception of the prestained protein molecular weight marker, which was ordered from New England Biolabs. Amicon YM-10 filtration products were acquired from Millipore (Billerica, MA). Y-Per yeast protein extraction reagent was obtained from Pierce (now a part of Fisher Thermo Scientific). Longlife Zymolyase was purchased from G-Biosciences (St. Louis, MO). Antibodies used for His_6 Western blotting procedures, monoclonal anti-polyhistidine clone His-I H1029, FLAG (M2 monoclonal), and anti-mouse IgG (alkaline phosphatase conjugate) A2429 were purchased from Sigma. Monoclonal antibodies for *myc* epitope, and the protein interleukin-12 were purchased from Invitrogen. Anti-sGP (MAI-21630) antibody from

Pierce was a kind gift of the Croyle Lab at the University of Texas at Austin. For Western blot detection, nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) were obtained from Promega. Porcine sequencing grade modified trypsin (V511A) was also obtained from Promega. Synthetic peptides were prepared by the Core Facilities of the Institute of Cellular and Molecular Biology at the University of Texas at Austin.

4.2.1.3 Plasmids Vectors and DNA Manipulation

Vectors pET24, pET28b(+), and pET32 vectors were purchased from Novagen (Madison, WI). Yeast expression vector pYES, pCR-Blunt and the pBad/Thio-Topo vectors were from Invitrogen (Carlsbad, CA), and yeast epitope tagging vector pESC-His was obtained from Stratagene (La Jolla, CA). Glycerol stocks of *E. coli* containing genes for the proteins TMTC2, TMTC3, and IL-12b on TOPO-4 plasmids were obtained from OPEN Biosystems (now a part of Thermo Fisher Scientific). DNA gel extraction and spin miniprep kits were obtained from Qiagen (Valencia, CA). DNA polymerase *pfu* was purchased from Stratagene. KOD DNA polymerase was purchased from Novagen (Madison, WI). Oligonucleotide primers for cloning were prepared by Invitrogen (Carlsbad, CA) or Integrated DNA Technologies (Coralville, IA).

4.2.1.4 Instrumentation

PCR reactions were carried out using an Eppendorf Mastercycler gradient thermal cycler from Brinkman-Eppendorf (Westbury, NY). For agarose gel electrophoresis, a mini-sub-cell GT from BioRad (Richmond, CA), powered by a Fisher FB-300 electrophoresis power source was used. After electrophoresis, DNA bands were

visualized using a Fisher FBTIV-08 UV transilluminator. Cell disruption was performed with a Fisher 550 Sonic Dismembrator equipped with a standard sonicator horn. Homogenization of cells was accomplished with a glass 20 mL Potter-Elvejhem homogenizer equipped with both loose (high-clearance) and tight-fitting (low clearance) pestles, which was purchased from Wheaton, (Millville, NJ). Thick-walled 4 mL polycarbonate centrifuge tubes (13 × 51 mm) were purchased from Beckman.

For SDS-page gel electrophoresis, a mini PROTEAN II vertical system from BioRad (Richmond, CA), equipped with a FB-300 (Fisher) or an EC-1000-90 (E-C apparatus corporation) electrophoresis power source was used. Mini-trans blot assembly apparatus used for Western blot, GelAir drying system for acrylamide gel preservation, and the necessary accessories and reagents were all products of BioRad. The nitrocellulose Hybond-C or Optitran BA-5 85 reinforced nitrocellulose membrane used for Western blot analysis was from Amersham Bioscences or Whatman (Dussel, Germany), respectively.

A Corning 240 pH meter equipped with an Accumet electrode (Fisher) was used to measure pH. Large scale centrifugation was performed using an Avanti J-25 unit or a J-E instrument from Beckman (Arlington Heights, IL). Microscale centrifugation was carried out using an Eppendorf 5415C from Brinkmann Instruments (Westbury, NY). Ultraviolet-visible spectra were recorded using a Beckman DU-650 spectrophotometer. In the case of Bradford assays, a Synergy HT plate reader by Biotek Instruments, equipped with a PC using KC4 software was used. HPLC separations were carried out using a Beckman 366 HPLC from Beckman Instruments (Fullerton, CA) equipped with an analytical C18 HPLC column from Varian (Lake Forest, CA). The ÄKTA FPLC was from Amersham Pharmacia Biosciences (now GE Healthcare). Sephadex 200 gel filtration FPLC columns were from Pharmacia (Uppsala, Sweden). DNA and biological

samples were dried using a Savant Speedvac SC100 equipped with a Savant refrigerated condensation trap RT100 and Gel Pump GP100. DNA concentrations were measured using a NanoDrop ND-1000 UV-Vis instrument from Thermo Fisher Scientific (formerly NanoDrop Technologies LLC). Ultracentrifuge Model TL-100, equipped with a Ti-100.4 rotor (both from Beckman) was used for microsomal preparations.

Sequencing of subclones was performed using a capillary-based AB 3700 DNA analyzer by the Core Facilities of the Institute of Cellular and Molecular Biology of the University of Texas at Austin. Mass spectra were obtained by the Mass Spectrometry Core Facility in the Department of Chemistry and Biochemistry using a Waters QTOF Premier mass spectrometer operating in electrospray mode, or MALDI-MS using a CHCA matrix by the College of Pharmacy of the University of Texas at Austin.

4.2.2 GENERAL PROCEDURES

The general methods and protocols for recombinant DNA manipulations were as described by Sambrook et al.[274] Protein concentrations were determined according to Bradford [272] using bovine serum albumin (BSA) as the standard. All protein purification operations were carried out at 4 ºC. The relative molecular mass and purity of enzyme samples were determined using SDS-polyacrylamide gel electrophoresis as described by Laemmli.[273]

4.2.3 EXPERIMENTAL PROCEDURES

4.2.3.1 Bacterial Strains and Growth Conditions

E. coli DH5α used for routine cloning experiments was maintained on Luria-Bertani (LB) broth or LB-agar plates. (1% NaCl, 1% tryptone, 0.5% yeast extract, pH 7.5, and in the case of plates, 2% agar). *E. coli* BL-21(DE3) and Rosetta strains used for routine gene expression experiments were maintained in Luria-Bertani (LB) broth or on LB-agar plates, supplemented with appropriate antibiotics (100 µg/mL ampicillin, 50 µg/mL kanamycin, and/or 35 µg/mL chloramphenicol). To liquid cultures, a volume of sterile 80% glycerol equal to 20% of the original culture volume was added and gently mixed to homogeneity. The mixture was aliquotted into sterile Eppendorf tubes, flash frozen in liquid N_2 and stored at -80 ºC until use. Glycerol stocks were used in a 1:20 dilution to inoculate fresh media for plasmid preparations. Competent cells were prepared as described in Chapter 2.

4.2.3.2 Eukaryotic Strains and Growth Conditions

S. cerevisiae InvSC1 was maintained on YPD-agar plates or cultured in YPD liquid media (1% yeast extract, 2% peptone, 2% dextrose, and in the case of plates, 2% agar). Yeast transformants were grown in SC dropout media (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.075% complete supplemental media without histidine or uracil, with the addition of 0.005% uracil or 0.01% histidine for metabolic selection, and in the case of plates 2% agar) Yeast strains were prepared for long term storage at -80 ºC by culturing overnight at 30 ºC with 200 RPM shaking in the appropriate media. Once the cultures reached OD_{600} > 1.0, a volume of sterile 80% glycerol equal to 20% of the original culture volume was added and gently mixed to homogeneity. The mixture was aliquotted into sterile Eppendorf tubes, flash frozen in

liquid N_2 and stored at -80 ºC until use. Glycerol stocks were used in a 1:20 dilution to inoculate fresh media for expression experiments.

4.2.3.3 PCR Reactions

Primers were prepared by dilution with TE buffer (50 mM Tris•HCl, pH 7.5, 1 mM EDTA, pH 8.0) to a concentration of 100 µM. This stock solution was further diluted to 10 or 20 µM with sterile water for KOD or *pfu* reactions respectively. pCR4-TOPO plasmid vectors containing each gene were obtained as *E. coli* DH5α glycerol stocks from OPEN Biosystems. Template DNA for TMTC2, TMTC3 and IL-12b genes was prepared by purifying the plasmid DNA from the *E. coli* DH5α strains using the Qiagen miniprep kit following the manufacturer's protocol. Polymerase chain reactions (PCR) primers and reaction programs vary for each gene and will be discussed in the appropriate sections of this chapter.

4.2.3.4 PCR Product Purification

Typically, PCR products were purified by agarose gel electrophoresis employing a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. The samples were diluted to 80% with 6X DNA loading dye (0.25% xylene cyanol, 0.25% bromophenol blue, 15% Ficoll 400) and electrophoresed at 100 V for 30 min. The DNA bands were visualized using an ultraviolet illumination source. DNA bands of the correct base pair length were excised from the gel with a scalpel and purified using the Qiagen gel extraction kit, according the manufacturer's protocols with the following modifications. Instead of heating at 55 ºC to melt the agarose, gel slices were heated at 65 ºC until the gel was fully

dissolved. DNA elution buffer EB was also heated to 65 ºC prior to DNA elution, and the elution volume was 30 µL.

4.2.3.5 Shuttle Vector Cloning into pGEM-T easy

PCR products were ligated into shuttle or expression vectors specific to each gene, as described in later sections of this chapter. For ligations into the pGEM-T easy vector, the DNA was first A-tailed. Briefly, 20 µL of purified DNA from the previous step was transferred to a 0.2 µL thin-walled PCR tube. To this was added 2.5 µL of 10X Taq polymerase buffer, 2 µL of a dNTP solution and 0.5 µL of Taq polymerase. The reaction was incubated without cycling for 10-30 min at 72 ºC. The A-tailed product was purified using the Qiagen gel purification kit following the modifications for purification of PCR products, and ligated into the shuttle vector pGEM-T easy. A typical ligation mixture included 2.5-3.0 µL of purified A-tailed PCR product, 0.2-0.5 µL of the pGEM-T easy vector, 2-2.5 µL of 5X T4 ligase buffer, and 0.5 µL of T4 ligase, for a total reaction volume of 6 µL.

The ligation mixture was incubated at room temperature or overnight at 4 ºC. The ligation mixture (3 µL) was mixed with competent *E. coli* DH5α (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, followed by incubation for 2 min on ice. LB broth (500 µL) was added to the transformed cells and the cells were pelleted by centrifugation at 12,000 *g* for 10 s, discarding the supernatant. Cells were resuspended in 50 µL fresh LB media and plated on LB-agar supplemented with 100 µg/mL ampicillin, 100 µL of 10 mM IPTG and 100 µM of 2% X-Gal in DMF. Positive transformants were selected by overnight incubation

at 37 ºC. Plates were subsequently incubated for 8-12 h at 4 ºC to allow development of the blue color. White colonies were selected for all future screening steps.

4.2.3.6 Shuttle Vector Cloning into pCR-Blunt

For cloning of PCR products into the shuttle vector pCR-Blunt, a typical ligation mixture included 3.5 µL of gel purified PCR product, 0.5 µL of pCR-Blunt vector, 2 µL of 5X T4 ligase buffer solution, 3.5 µL of sterile water, and 0.5 µL of T4 ligase, for a total reaction volume of 10 µL. The ligation mixture was incubated at room temperature for 30 min-1 h, or overnight at 4 ºC. The ligation mixture (3-5 µL) was mixed with competent *E. coli* DH5α (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, followed by incubation for 2 min on ice. LB broth (500 µL) was added to the transformed cells and incubated for 30-45 min at 37 ºC, with 200 RPM shaking. The cells were plated on LB-agar plates containing 50 µg/mL kanamycin. Positive transformants were selected by incubation overnight at 37 ºC.

4.2.3.7 DNA Extraction by Alkaline Lysis

Colonies appearing after overnight incubation were screened for the presence of plasmid containing PCR product using a modified alkaline lysis procedure.[277] Briefly, 2-5 mL cultures of each positive transformant were grown overnight at 37 ºC in LB media in the presence of selective antibiotic. Culture volume was dependent on the copy number of the plasmid. Cultures were centrifuged at 12,000 *g* for 30 s, the supernatant was removed and the cells were resuspended in 100 µL of cold resuspension solution (50 mM glucose, 25 mM Tris•HCl, pH 8.0, 10 mM EDTA, 100 µg/mL RNase). Cells were

disrupted with the addition of 200 µL fresh lysis solution (200 mM NaOH, 1% SDS), and inverted five times to mix. The lysis reaction was quenched with the addition of 150 µL of neutralization solution (3 M potassium acetate, 11.5% acetic acid). Reactions were mixed to ensure neutralization and incubated at 4 ºC for 5 min, followed by centrifugation for 5 min at 16,100 *g* at 4 ºC to precipitate cellular debris. Each supernatant containing plasmid DNA was transferred to a fresh tube and the DNA was precipitated by the addition of 2.5 volumes of 100% ethanol or one volume of isopropanol and 1/10 volume of 3 M sodium acetate, pH 5.2. Each solution was mixed to ensure homogeneity and centrifuged at 16,100 *g* for 15 min at 4 ºC. The supernatant from each sample was removed by aspiration, taking care not to disturb the DNA pellet. Each DNA pellet was washed with 70% ethanol and centrifuged again at 16,100 *g* for 5 min. The final ethanol wash was removed and each pellet was dried using a centrifugal vacuum apparatus for 2 min at room temperature to remove any remaining ethanol. Each pellet containing plasmid DNA was resuspended in 30 µL of 50 mM Tris•HCl (pH 8.5).

4.2.3.8 DNA Analytical Restriction Digestion and Fragment Analysis

Plasmid DNA obtained from the alkaline lysis preparation was subjected to enzymatic digestion with the appropriate restriction enzymes. A typical reaction mixture contained 5 µL of DNA, 2 µL of 10X NEB reaction buffer, 0.2 µL of each enzyme, and 13 µL of sterile water, for a final volume of 20 µL. In the case of shuttle vectors pGEM-T easy or pCR-Blunt, a typical reaction mixture contained 3 µL of DNA, 4 µL of 10X *EcoR*I buffer, 0.5 µL of *EcoR*I, and 13 µL of sterile water, for a final volume of 20 µL. Each reaction was incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis.

Samples were electrophoresed for 30 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. For digestion products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragments. Electrophoretic results as visualized on agarose gels were used to determine the presence of the DNA insert of the appropriate length.

Colonies containing the vector with the ligated PCR product were grown at 37 ºC in LB media supplemented with selective antibiotic. After overnight incubation, cells were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the pelleted bacteria using the Qiagen miniprep kit, according to the manufacturer's protocols with the following adjustments. The optional 500 µL PB wash was always performed to remove additional contaminants, and elution buffer EB was heated to 65 ºC before DNA elution.

4.2.3.9 DNA Sequence Analysis

The purified plasmid was sequenced to confirm the identity of the PCR product. For these genes, sequencing reactions using external and internal primers were submitted to ensure full sequence coverage. Specific primers used for each sequencing reaction are provided in later sections of this chapter. PHRED sequences obtained from the DNA facility were aligned against the known gene sequence. Contigs consisting of results from DNA sequencing reactions were assembled using the Vector NTI software program and compared to the original gene sequence. Both forward and reverse sequences were examined to determine that the correct gene sequence was obtained. When necessary, point mutations in the amplified sequences were compared to known Single Nucleotide

Polymorphisms (SNPs) to determine if any of these is a result of mutations in the DNA template.

4.2.3.10 Secondary Cloning into Expression Vectors

When shuttle vectors were used, the PCR insert was isolated after sequencing and ligated into a plasmid expression vector. The general protocol used to generate expression constructs is described below.

STEP 1: GROWTH OF E. COLI CELLS. One colony of *E. coli* DH5α containing the expression vector or the gene of interest in a shuttle vector was used to inoculate 5 mL of LB media supplemented with the appropriate selective antibiotic. Cultures were grown overnight at 37 ºC with 200 RPM shaking and the plasmids were isolated using the Qiagen miniprep kit following the manufacturer's protocols with additions and alterations as previously detailed.

STEP 2: SHUTTLE VECTOR DIGESTION. Plasmid DNA obtained from this preparation was subjected to enzymatic digestion overnight with the appropriate enzymes. A typical preparatory digestion reaction contained: 30 µL of DNA, 4 µL of NEB 10X buffer, 2 µL of each enzyme and 2 µL of sterile water (or 1.6 µL of sterile water and 0.4 µL of 100X BSA) for a final reaction volume of 40 µL. For yeast expression constructs (*tmtc*2- and *tmtc*3-pCR-Blunt and *il-12b*-pGEM-T easy), a typical reaction mixture consisted of 10 µL DNA, 5 µL of 10X NEB buffer, 35 µL of sterile water and 1 µL of each enzyme. Reactions were incubated overnight at 37 ºC. Digestion reactions were subjected to agarose gel electrophoresis for 30-40 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide to separate the gene of interest from the shuttle vector. For digestion products greater than 2 kB, the duration of the electrophoretic separation was

increased to allow greater resolution of digested vector and the resulting DNA fragment. The gel band containing the insert of appropriate size was excised with a clean, sharp scalpel and purified using the Qiagen gel purification kit, following the manufacturer's protocols with the additions and alterations as previously described.

Step 3: Expression Vector Purification and Digestion. Expression vectors were prepared by enzymatic digestion with the appropriate enzymes overnight at 37 ºC, followed by the addition of 0.1 µL calf intestinal alkaline phosphatase (CIP). The reaction was incubated for 1 h at 37 ºC. When NEB buffer 2 was not used in the digestion reaction, 0.25 volumes of 1 M Tris, pH 8 was added to the reaction. To purify digested DNA, the Qiagen gel purification kit was employed according to manufacturer's instructions for purifying PCR products.

Step 4: DNA Ligation and Transformation into E. coli. A Nanodrop instrument was used to determine the concentration of both vector and gene insert solutions. Ligations were performed at room temperature for 1-2 h or overnight at 4 ºC, with a typical reaction mixture consisting of 90 fmol ends of gene to be inserted, 30 fmol ends of vector, 2 µL of 5X ligation buffer, 0.5 µL of T4 ligase, and sterile water to a final volume of 10 µL. A total of 5 µL of the ligation reaction was incubated with 50 µL of competent *E. coli* DH5α on ice for 30 min. The transformation procedure is identical to that previously described, except when ampicillin was used as the selective antibiotic, the recovery incubation was omitted and the transformation was plated directly onto LB-agar plates in the presence of 100 µg/mL ampicillin.

Step 5: Ligation and Selection Analysis. After overnight incubation at 37 ºC, colonies appearing on the selection plates were screened for successful ligation by alkaline lysis and enzymatic digestion following the method described above with the following adjustments.[277] Digestion was carried out using the restriction enzymes specific

for each construct, as indicated in the following chapter. A typical reaction mixture contained 5 µL of DNA, 2 µL of 10X NEB reaction buffer, 0.2 µL of each enzyme, and 13 µL of sterile water, for a final volume of 20 µL. Each reaction was incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis. Samples were electrophoresed for 30 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. For digestion products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragment.

Electrophoretic results, as visualized on agarose gels, were used to determine the presence of the DNA insert of the appropriate length. Colonies containing the vector with the successfully ligated gene of interest were grown at 37 ºC in LB media supplemented with the appropriate selective antibiotic. After overnight incubation, cells were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the pelleted bacteria using the Qiagen miniprep kit, following the manufacturer's protocols with the adjustments as described previously. The newly isolated gene-containing plasmids were used to transform *E. coli* BL-21, BL-21*, BL-21(DE3), BL-21(DE3)pLysS and BL-21 RosettaII (DE3) heterologous expression hosts or *S. cerevisiae* competent cells.

4.2.3.11 SDS-PAGE

The relative molecular mass and purity of enzyme samples were determined using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli.[273] Proteins were electrophoresed using a discontinuous buffer system containing 25 mM Tris•HCl, 192 mM glycine and 0.1% SDS (pH 8.3).[273] Gels were cast using 75 mm plates, and in most cases 6 or 8% acrylamide was used for the separatory

gel, while the stacking gel was 4% acrylamide. Protein samples were diluted 2X with SDS-loading dye (62.5 mM Tris•HCl buffer pH 6.8, containing 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.0025% bromophenol blue), and heated to 100 ºC for 5 min. Samples were centrifuged briefly to collect protein at the bottom of the tubes. The protein gels were electrophoresed using a setting of 30 mA per gel, allowing the gel to run until the loading dye eluted from the bottom of the gel. Gels were removed from the glass plates, and the stacking gel was removed and discarded. A corner of each gel was removed to preserve orientation of the gel samples. Gels were stained with Coomassie blue (2.5 g/L of Coomassie Brilliant Blue G-250 in acetic acid:water:methanol (1:4:5) by volume) until gels were dark blue.[278] Gels were removed from staining solution, rinsed in ddH_2O, and introduced to destaining solution (ethanol:acetic acid:water in a ratio of 4:5:41 by volume).[278] Protein concentrations were determined according to Bradford using BSA as the standard, and the dye reagent from BioRad. [272]

4.2.3.12 Western Blot Procedure

Western blotting was used to detect the presence of IL-12b or FLAG and *c-myc* epitopes, or polyhistidine tags on heterologously expressed and purified proteins.[279] SDS-PAGE procedures were performed in duplicate for each protein sample. The gels were processed as described previously until the staining step, at which point gels were assembled into the transfer cassettes. The gels were transblotted onto the nitrocellulose membrane at 100 V for 1 h, using a running buffer of 25 mM Tris base, 192 mM glycine, 20% v/v methanol (pH 8.3). Cassette assembly and electrophoresis were carried out according to the directions provided by BioRad. After transblotting electrophoresis, the membrane was carefully removed from the transfer cassette and the presence of the

prestained marker on the blot was visually assessed to ensure transfer was complete. The membrane was transferred to a small plastic box and immersed in 5% non-fat dry milk in Tris buffered saline and Tween (TBST) solution (20 mM Tris•HCl, 150 mM NaCl, 0.5% v/v Tween® 20, pH 7.5). The membrane was blocked in this solution for 1 h, with agitation, overnight at 4 ºC or at room temperature for 1 h. The solution was removed and the membrane was rinsed thrice in TBST to remove excess blocking solution. The membrane was immersed in primary antibody solution (15 mL TBST, 1:30,000 dilution of monoclonal anti-polyhistidine clone His-1, or a 1:5000 dilution of α-IL-12b, α-myc, or α-FLAG. All antibody solutions contained a 1:1000 dilution of 20% v/v NaN_3.) and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was removed from the membrane and the membrane was washed thrice with TBST to remove any excess primary antibody solution remaining on the membrane. The membrane was subsequently incubated in the secondary antibody solution (15 mL TBST, 1:30,000 dilution of anti-mouse IgG [Fc specific] alkaline phosphatase conjugate, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was removed and the membrane was washed thrice with TBST to remove any excess secondary antibody solution remaining on the membrane. To visualize proteins, the membrane was immersed in 5 mL color development solution (100 mM Tris•HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) with the addition of 66 µL NBT and 33 µL BCIP. The membrane was incubated in the dark for 2-5 min or until purple bands appeared on the blot. Development was stopped by rinsing the membrane with ddH_2O. Membranes were dried and scanned within 24 h.

4.2.3.13 In-Gel Tryptic Digestion of Proteins

Proteins were digested in acrylamide gels following the method of Shevchenko.[328] Briefly, major protein bands were excised from Coomassie stained gel with a fresh razor blade. Gel slices were transferred to sterile eppendorf tubes and were destained overnight at room temperature with slight agitation in fresh destaining solution. After most of the dye was removed from the gel bands, destaining solution was removed and replaced with 500 µL of fresh destaining solution, and destaining continued for another 2 to 3 h. The destaining solution was removed and discarded, and the gel slices were dehydrated in 300 µL of HPLC grade acetonitrile. This step was repeated as necessary until the gel slices were an opaque white color. The acetonitrile was removed and discarded. The gel slices were subjected to evaporation using vacuum centrifugation for 2-3 min to remove any remaining acetonitrile and treated with fresh 10 mM DTT (100 µL) for 1 h at room temperature to reduce the proteins in the gel. The sample was centrifuged briefly and the DTT solution was removed. The gel slices were incubated at room temperature for 1 h with fresh 50 mM iodoacetamide solution (100 µL). The sample was briefly centrifuged and the iodoacetamide solution was removed. The gel slices were washed for 10 min with 100 mM ammonium bicarbonate (200 µL), the sample was centrifuged briefly and the ammonium bicarbonate solution was removed. Gel slices were dehydrated in HPLC grade acetonitrile (300 µL). This step was repeated as necessary until the gel slices turned white, after which the acetonitrile was left on the gel slices for 5 min. The acetonitrile was removed and discarded. This step was repeated for two washes of 5 min each, and the gel was rehydrated in 100 mM ammonium bicarbonate (200 µL) for 10 min. The sample was centrifuged briefly and the ammonium bicarbonate solution was removed. Gel slices were dehydrated in HPLC grade acetonitrile (300 µL). This step was repeated twice as above, for a total of two 5 min washes. The gel slices were dried by evaporation

using the vacuum centrifuge for 2-3 min to remove any remaining acetonitrile. Trypsin solution (20 ng/μL) was prepared by adding 1 mL of 50 mM ammonium bicarbonate to 20 μg enzyme on ice. Trypsin solution (50- 100 μL) was added to cover the gel slices completely, and this mixture was incubated on ice for 10-15 min, until the gel pieces appeared swollen. Ammonium bicarbonate (20 μL of a 50 mM solution) was added to the solution and the digestion was incubated overnight at 37 ºC. The next morning, the extract was centrifuged to collect condensate and the digestion solution was removed with a gel-loading pipette tip, ensuring exclusion of any gel pieces. Protein was further extracted from the gel slices by adding 50% acetonitrile/50% water with 5% formic acid (75 μL) and incubating at room temperature for 10 min. The sample was centrifuged and the resulting solution transferred to a clean eppendorf tube. This step was repeated for a total organic extraction volume of 150 μL.

4.2.4 First Attempt to Identify C-Mannosyltransferase: Microsomal Crosslinking Assays

4.2.4.1 Preparation of Rat Liver Microsomes

Microsomes were prepared by sucrose gradient centrifugation according to the method of Graham.[329] Adult female rats were sacrificed by decapitation and bled through the neck to exsanguinate. Livers were immediately removed and transferred to a chilled, pre-weighed beaker containing homogenization medium (250 mM sucrose, 5 mM Tris•HCl pH 7.4, 1 mM $MgCl_2$) Livers were weighed and rinsed twice in homogenization buffer to remove excess fluids. All remaining steps were carried out between 0-4 ºC. Tissue was chopped finely and mixed with homogenization media (0.25 g liver/mL). Tissues were homogenized in a Potter-Elvehjem homogenizer and the cell debris pelleted

by centrifugation at 1,000 *g* for 10 min at 4 ºC. Further cellular debris was removed by centrifugation (4 ºC) at 3,000 *g* for 10 min and 10,000 *g* for 20 min. The resulting supernatant was layered on top of a density barrier comprised of 1.5 mL of Solution I (0.6 M sucrose, 15 mM CsCl, 5 mM Tris•HCl, pH 8.0) and 3 mL of Solution II (1.3 M sucrose, 15 mM CsCl, 5 mM Tris•HCl, pH 8.0). The discontinuous gradient was centrifuged for 90 min at 100,000 *g* using a Beckman Ti-70.1 rotor. The pellet, which contains the rough microsomes, was washed with 50 mM potassium phosphate buffer, pH 7.5 and resuspended in 2 mL of 20 mM HEPES-NaOH pH 7.2, containing 110 mM potassium acetate and 2 mM magnesium acetate (KMH buffer) and the protease inhibitor cocktail (leupeptin, 2.4 μg/mL; antipain, 4 μg/mL; lima bean trypsin inhibitor, 40 μg/mL; benzamidine, 20 μg/mL; AEBSF HCl, 10 μg/mL; aprotinin, 2 μg/mL; chymotrypsin, 10 μg/mL; pepstatin A, 5 μg/mL). Microsomes were aliquotted, flash frozen in liquid nitrogen, and stored at -80 ºC. The protein concentration was determined by the method of Bradford.[272]

4.2.4.2 5-Hydroxytryptophan Crosslinking Assay

In an attempt to isolate the *C*-mannosyltransferase, an assay was designed to crosslink the CMT to an engineered substrate. A small peptide substrate N-KPPQFAW*AQWFE-C (designed after Doucey) was prepared by solid phase synthesis.[315] An intrinsic crosslinker was engineered by substituting the tryptophan at the position of mannosylation with 5-hydroxytryptophan, as indicated by W* above. The peptide (1 μmol and 5 μmol), linked to agarose beads, was incubated with 200 μL of microsomal protein in the presence of 0.6 molar equivalents of sodium periodate. To increase protein solubility and to permeabilize microsomes, 0.2% Triton-X 100 was

added to the reaction. Each reaction was brought to 1 mL final volume with KMH buffer and incubated for 4 h at room temperature. Sodium borohydride was added in excess to quench the reaction. The supernatant was removed and the beads washed 3 times with KMH buffer to remove protein that was not crosslinked. Beads were resuspended in 100 μL of KMH buffer and 20 μL of the resulting solution was removed and added to an equal volume of 2X SDS loading buffer in the presence of β-mercaptoethanol. These samples were boiled for 30 min and 20 μL of each sample was electrophoresed on a denaturing SDS-PAGE gel. The gel was stained with Coomassie blue stain to visualize protein bands.

4.2.4.3 CMT Crosslinked Protein In-Gel Digestion

The major protein bands were excised from the Coomassie stained gel with a fresh razor blade. The gel slices were processed according to the method of Shevchenko, as described previously.[328] The samples were concentrated to 25 μL using the vacuum centrifuge and submitted for MS analysis.

4.2.5 SECOND ATTEMPT TO IDENTIFY C-MANNOSYLTRANSFERASE: IN SILICO GENE ANALYSIS

Using gi 34190043 (*H. sapiens* protein *O*-mannosyltransferase 1, POMT-1) as our query, this protein sequence was subjected to pblast analysis using the NCBI "Blast" website. Among all protein sequences in the database, 23 proteins were identified as having some homology to POMT-1. These protein sequences were analyzed manually by comparison to POMT-1, and that protein TMTC2 was identified as a possible *C*-mannosyltransferase.

4.2.6 CLONING OF TMTC2 FOR E. COLI EXPRESSION

4.2.6.1 PCR Reactions

The *tmtc2* gene was amplified using the pCR-TOPO4 vector as template DNA. Primers were designed for direct cloning into the pBad/Thio-Topo vector. The primer sequences (5′ start primer, 5′-ATGATTGCAGAGTTGGTGAGCAGCGCTCTG-3′, and 3′ halt primer, 5′-GGTCTTAGAAGTCTTTAAGCCTTGTTTTTCCATGATGTTCCAC-AG-3′) were designed to remove the stop codon from the amplified gene such that a C-terminal V5 epitope and a His_6-tag were introduced upon gene expression.

The primers were obtained as lyophilized powder and were prepared as described previously. The PCR reactions consisted of the following: 72 µL of sterile water, 10 µL of 10X *pfu* buffer, 10 µL of a 20 mM dNTP solution, 2.5 µL of each primer (5′ and 3′) at 20 µM, 1 µL of template DNA, 1.8 µL of *pfu* polymerase, and 0.2 µL of Taq polymerase for a 100 µL total reaction volume in an 0.5 mL thin walled PCR tube. The PCR reactions were carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial denaturing step of 2 min at 95 ºC. The program consisted of a 30 s denaturing step at 95 ºC, an annealing step at 58.8 ºC for 30 s, and an elongation step of 3 min (1 min/kb) at 72 ºC. This cycle was repeated thirty times, followed by a 10 min extension at 72 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

4.2.6.2 pBAD-TMTC2 Cloning

The PCR products were purified by agarose gel electrophoresis as described in section 4.2.3.4 and ligated into the pBad/Thio-Topo vector. A typical ligation mixture

included 4 µL of gel purified PCR product, 1 µL of NaCl solution, and 1 µL of TOPO vector. The ligation mixture was incubated at room temperature for 30 min. The ligation mixture (2 µL) was mixed with competent *E. coli* DH5α (50 µL) and transformed following the procedure described in section 4.2.3.5. Positive transformants were selected after incubation overnight at 37 ºC on LB-agar plates supplemented with 100 µg/mL ampicillin.

Colonies appearing after overnight incubation were screened for the presence of plasmid containing the PCR product using a modified alkaline lysis procedure as described in section 4.2.3.7.[277] Plasmid DNA obtained from this preparation was subjected to enzymatic digestion with *Nco*I and *Pme*I followed by agarose gel electrophoresis.

Colonies containing the vector with the ligated PCR product were grown overnight at 37 ºC in LB media supplemented with 100 µg/mL ampicillin. Plasmid DNA was extracted from the bacterial pellet using the Qiagen miniprep kit, following the manufacturer's protocols with the adjustments described previously in section 4.2.3.7.The purified plasmid was sequenced to confirm the identity of the PCR product. For the TMTC2 gene, five sequencing reactions, using two external primers (Trx forward and pBAD reverse) from the cloning vector and three internal primers were submitted to ensure full sequence coverage. Sequencing was analyzed as described in section 4.2.3.9 and colonies containing the gene of interest having the correct sequence were cultured as described above. Plasmid pBAD-TMTC2 was isolated using the Qiagen kit with the adjusted protocol as previously described and used to transform *E. coli* heterologous expression hosts.

4.2.6.3 pET-TMTC2 Cloning

To create the pET32a construct, a 5 mL culture of *E. coli* DH5α cells containing pCR-TOPO4-TMTC2 was grown overnight at 37 ºC in LB media supplemented with 50 µg/mL kanamycin. After overnight growth, the plasmid was isolated using the Qiagen miniprep kit following the manufacturer's protocols with adjustments as previously described. Plasmid DNA obtained from this preparation was subjected to enzymatic digestion overnight with *Nco*I and *Pme*I following the standard procedure provided in section 4.2.3.10. After electrophoresis, the DNA band of the correct size was extracted from the gel following the methods described earlier in this chapter. Expression vectors were prepared by enzymatic digestion with the appropriate enzymes (*Nco*I and *EcoR*V), and purified according to the protocols described in section 4.2.3.10. Ligations followed the procedures given earlier in the chapter and the transformation procedure is identical to that previously described (see section 4.2.3.10), and the transformation was plated directly onto LB-agar plates containing 100 µg/mL ampicillin.

After overnight incubation at 37 ºC, colonies appearing on the selection plates were screened for successful ligation by alkaline lysis and enzymatic digestion according to the methods described above with the following adjustments. Digestion was carried out using the restriction enzymes *Nco*I and *Xho*I, Positive colonies were amplified and gene-containing plasmids purified for transformation into *E. coli* BL-21(DE3). This construct was created such that the *tmtc2* gene would be expressed as an *N*-terminal thioredoxin fusion protein with a His_6-tag.

4.2.7 *E. coli* Transformations and Attempts to Express the *tmtc2* Gene

STEP 1: GROWTH OF E. COLI CELLS. Plasmid constructs pET-TMTC2 and pBad-TMTC2 were used to transform competent *E. coli* BL-21(DE3) Rosetta II cells following

the previously described protocol. Selection was performed by plating the recombinants on LB-agar supplemented with 100 µg/mL ampicillin and incubated overnight at 37 ºC. Small-scale cultures were inoculated from colonies grown on the selective plate. Cultures of LB broth (20 mL) supplemented with 100 µg/mL ampicillin were inoculated with single colonies and grown overnight with 200 RPM shaking. These starter cultures were used to inoculate cultures grown in different media at various temperatures with different growth times. Media used include Luria-Bertani (LB), Terrific Broth (TB) (1.2% tryptone, 2.4% yeast extract, 0.4% glycerol, and 100 mL of KP_i buffer (0.17 M KH_2PO_4/ 0.72 M K_2HPO_4)) and M9ZB media (0.5% NaCl, 0.1% NH_4Cl, 0.3% KH_2PO_4, 0.6% Na_2HPO_4, 1% N-Z-Amine A, 0.4% glucose, 1 mM $MgSO_4$).[330] Cultures were induced at varying OD_{600} as indicated in the table below and the inducer varied based on the construct used. Each of the cultures was centrifuged at 4 ºC to pellet the cells. The cell pellets were stored at -80 °C for protein purification. A total of 56 trials were conducted to determine the best expression conditions for TMTC2 in *E. coli*, (Table 4.2).

4.2 Expression conditions for TMTC2 in *E. coli*

Media	Construct	Growth Temp (ºC)	Induction Temp (ºC)	Time (h)	OD_{600}	Induction (mM IPTG/%L-Ara)	Preparation
LB	pET32a	37	18	3	N/A	0 mM	I
TB	pET32a	37	18	3	N/A	0 mM	I
LB	pET32a	37	18	3	0.8	1.0 mM	I
TB	pET32a	37	18	3	0.8	1.0 mM	I
LB	pET32a	37	18	12	N/A	0 mM	I
TB	pET32a	37	18	12	N/A	0 mM	I
LB	pET32a	37	18	12	0.8	1.0 mM	I
TB	pET32a	37	18	12	0.8	1.0 mM	I
LB	pET32a	30	16	3	N/A	0 mM	II
TB	pET32a	30	16	3	N/A	0 mM	II
LB	pET32a	30	16	3	0.8	1.0 mM	II
TB	pET32a	30	16	3	0.8	1.0 mM	II

LB	pET32a	30	16	12	N/A	0 mM	II
TB	pET32a	30	16	12	N/A	0 mM	II
LB	pET32a	30	16	12	0.8	1.0 mM	II
TB	pET32a	30	16	12	0.8	1.0 mM	II
LB	pET32a	37	14	3	N/A	0 mM	I
LB	pET32a	37	14	12	N/A	0 mM	I
LB	pET32a	37	14	3	1	1.0 mM	I
LB	pET32a	37	14	12	1	1.0 mM	I
LB	TrxTOPO	37	16	12	0.6-0.7	0	I
LB	TrxTOPO	37	16	12	0.6-0.7	0.000002%	I
LB	TrxTOPO	37	16	12	0.6-0.7	0.00002%	I
LB	TrxTOPO	37	16	12	0.6-0.7	0.0002%	I
LB	TrxTOPO	37	16	12	0.6-0.7	0.002%	I
LB	TrxTOPO	37	16	12	0.6-0.7	0.2%	I
LB	TrxTOPO	37	16	12	0.6-0.7	2%	I
LB	TrxTOPO	30	16	12	0.8	2%	I
TB	TrxTOPO	30	16	12	0.9	2%	I
M9ZB	TrxTOPO	30	16	12	0.9	2%	I
LB	TrxTOPO	30	16	12	0.8	0%	I
TB	TrxTOPO	30	16	12	0.9	0%	I
M9ZB	TrxTOPO	30	16	12	0.9	0%	I
LB	TrxTOPO	30	16	3	0.6-0.7	2%	I
TB	TrxTOPO	30	16	3	0.9	2%	I
M9ZB	TrxTOPO	30	16	3	0.9	2%	I
LB	TrxTOPO	30	16	3	0.8	0%	I
TB	TrxTOPO	30	16	3	0.6-0.7	0%	I
M9ZB	TrxTOPO	30	16	3	0.9	0%	I
LB	TrxTOPO	30	16	12	0.8	2%	II
TB	TrxTOPO	30	16	12	0.9	2%	II
M9ZB	TrxTOPO	30	16	12	0.9	2%	II
LB	TrxTOPO	30	16	12	0.8	0%	II
TB	TrxTOPO	30	16	12	0.9	0%	II
M9ZB	TrxTOPO	30	16	12	0.9	0%	II
LB	TrxTOPO	30	16	3	0.8	2%	II
TB	TrxTOPO	30	16	3	0.9	2%	II
M9ZB	TrxTOPO	30	16	3	0.9	2%	II
LB	TrxTOPO	30	16	3	0.8	0%	II
TB	TrxTOPO	30	16	3	0.9	0%	II
M9ZB	TrxTOPO	30	16	3	0.9	0%	II
LB	TrxTOPO	37	37	3	~1	2%	II
M9ZB	TrxTOPO	37	37	3	~1	2%	II
LB	TrxTOPO	37	37	3	~1	0%	II
M9ZB	TrxTOPO	37	37	3	~1	0%	II

4.2.8 E. COLI PROTEIN PURIFICATION METHODS FOR TMTC2- ISOLATING TRANSMEMBRANE PROTEIN

STEP 2A: CRUDE EXTRACT PREPARATION. Preparation I follows a method reported by the Imperiali group.[109] Cells were harvested by centrifugation at 4,500 *g* for 15 min at 4 ºC and frozen overnight at -80 ºC.

STEP 3A: METHOD I. All subsequent procedures were carried out at 4 ºC or on ice. Cells were resuspended in 1/20 volume of 50 mM Tris-acetate, pH 8.0, 1 mM EDTA buffer and lysed by sonication. Cells were centrifuged at 5,700 *g*. The supernatant fractions from these centrifuged samples were subjected to ultracentrifugation at 142,400 *g* for 1 h. The resulting pellet was resuspended in Tris-acetate buffer with 30% w/v glycerol and stored at -20 ºC.

STEP 2B: CRUDE EXTRACT PREPARATION. The preparation II procedure was reported by the Schutzbach group. [331] Volumes given in this protocol are specific for a 1 L preparation. Cultures were harvested by centrifugation at 6,700 *g* at 4 ºC for 15 min.

STEP 3B: METHOD II. All subsequent procedures were carried out at 4 ºC or on ice. Cells were washed with 300 mL of water followed by a wash using 100 mL of buffer B (25 mM sodium phosphate, pH 8.0, 5 mM $MgCl_2$, and 0.2% β-mercaptoethanol). The cell pellet was frozen and stored at -80 ºC overnight. The pellet was thawed, resuspended in 10 mL of buffer B, and the cells were ruptured by sonication at a power setting of 4.5 for 3 × 30 s with a 1 min interval to allow cooling. The particulate (pellet) fraction obtained by centrifugation at 39,000 *g* for 20 min, was washed twice with 25 mL of buffer B and resuspended in 5 mL of the same buffer. The suspension was diluted with 17.5 mL of buffer C, (10 mM sodium phosphate, pH 8.0, containing 0.5 mM EDTA, 10% glycerol, 0.1% Nonidet P-40, and 1.0 mM dithiothreitol). Nonidet P-40 (2.5 mL of a 10% solution) was added, and the mixture was centrifuged at 39,000 *g* for 20 min. The

supernatant was removed and retained as “fraction C”. Additional enzyme was solubilized by the addition of 20 mL of buffer D (0.1 M Tris acetate, pH 7.5, containing 10% (v/v) glycerol, 0.2% β-mercaptoethanol, and 0.1% (w/v) sodium dodecyl sulfate) to the precipitate from the previous step. The precipitate was dispersed by vigorous trituration followed by the addition of 2.5 mL of 10% Nonidet P-40. The solution was centrifuged as above, and the supernatant was removed and retained as “fraction D”. This procedure was repeated once more to yield a third solubilized fraction, which was combined with “fraction D” The pellet was retained as “insoluble fraction”. All fractions were stored at -20 ºC.

STEP 4: SDS-PAGE AND WESTERN BLOTTING. Membrane fractions from Method I and Fractions C and D from Method II were analyzed using SDS-PAGE denaturing conditions as described in section 4.2.3.11, and electrophoresis was performed in duplicate. One gel was subjected to Coomassie staining, and the other was transferred to a nitrocellulose membrane for Western blotting. Membranes were probed using Anti-V5 and Anti-thioredoxin primary antibodies for pBad-TMTC2 constructs. Anti-thioredoxin and Anti-His primary antibodies were used to detect full-length transcript for pET-TMTC2 constructs.

4.2.9 CLONING OF IL-12B FOR E. COLI EXPRESSION

4.2.9.1 PCR Reactions and Product Purification

The IL-12b gene was amplified using the IL-12b-pCR-TOPO4 vector as template DNA. Primers for *il-12b* expression were designed for cloning into the pET-32a vector. The primer sequences incorporated *Nde*I and *Hind*III restriction sites in start and halt

primers respectively, (5′ start primer, 5′- GTCGATC*CATATG*TGTCACCAGC-AGTTG-3′, and 3′ the halt primer, 5′-ATCG*AAGCTT*CTAACTGCAGGGCAC-3′) and were designed to generate a construct such that an *N*-terminal thioredoxin and His_6-tag would be introduced upon gene expression.

Template and DNA primers were obtained and prepared for PCR as described previously. The PCR reactions consisted of the following: 32.8 µL of sterile water, 5 µL of 10X KOD buffer, 5 µL of a 2.0 mM dNTP solution, 3.0 µL of a 25 mM $MgSO_4$ solution, 1.5 µL of a 10 µM solution of each primer (5′ and 3′), 0.2 µL of template DNA, and 1.0 µL of KOD polymerase, for a 50 µL total reaction volume in an 0.2 mL thin-walled PCR tube. PCR was carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial denaturing step of 2 min at 95 ºC. The program consisted of a 30 s denaturing step at 95 ºC, an annealing step at 55.9 ºC for 10 s, an elongation step of 20 s (20 s/kb) at 70 ºC. This cycle was repeated thirty times, and the program ended with a 10 min extension at 70 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

The PCR products were purified by agarose gel electrophoresis as described in section 4.2.3.4, with the elution volume reduced to 20 µL.

4.2.9.2 IL-12b Shuttle Vector Cloning

PCR products were A-tailed and ligated into the shuttle vector pGEM- T easy for blue-white screening as specified in section 4.2.3.5. After this procedure, the purified plasmid was sequenced to confirm that the sequence of the PCR product is correct. For

the *il-12b* gene, two sequencing reactions using the M13 external primers were performed to ensure full sequence coverage.

4.2.9.3 Cloning of il-12b into pET32a Plasmid for E. coli Expression

The pET32-il-12b expression construct was generated following protocols detailed in section 4.2.3.10. Specific details are as follows. All liquid and solid media was supplemented with 100 µg/mL ampicillin for selection. Restriction enzymes used for digestion were *Nde*I and *Hind*III enzymes. Culture volumes for pET vector amplification or isolation were 10 mL. Cultures and plate were incubated for 8-10 h only. The plasmid generated via this method was pET32a/IL-12b for expression of IL-12b in *E. coli* as an *N*-terminal thioredoxin (Trx) fusion protein. Positive colonies were amplified and gene-containing plasmids purified for transformation into *E. coli* BL-21 and BL-21*.

4.2.10 E. COLI TRANSFORMATIONS AND IL-12B GENE EXPRESSION TRIALS

The pET32a/IL-12b plasmid construct was transformed into competent *E. coli* BL-21 and BL-21* cells following the protocol described previously and selected by plating on LB-ampicillin plates (100 µg/mL) after overnight incubation at 37 ºC. Colonies that grew were used to inoculate small-scale (20 mL) cultures and trials were conducted to determine the best expression conditions for IL-12b in *E. coli* according to the table below.

STEP 1: GROWTH OF E. COLI CELLS. Cultures of LB broth (6 × 20 mL) supplemented with 100 µg/mL ampicillin were inoculated with single colonies and grown overnight with 200 RPM shaking. Two cultures were grown at 30 °C, two at 37 °C, and two at 15 ºC.

4.3 Expression conditions for IL-12b in *E. coli*

Media	Cell Line	Growth Temp ºC	Expression Temp ºC	Time (h)	OD_{600}	Induction (mM IPTG)	Preparation
LB	BL-21	37	N/A	12	N/A	0 mM	Crude soluble
LB	BL-21	30	N/A	16	N/A	0 mM	Crude soluble
LB	BL-21	15	N/A	30	N/A	0 mM	Crude soluble
LB	BL-21*	37	N/A	12	N/A	0 mM	Crude soluble
LB	BL-21*	30	N/A	16	N/A	0 mM	Crude soluble
LB	BL-21*	15	N/A	30	N/A	0 mM	Crude soluble
LB	BL-21*	37	N/A	12	0.5	0.1 mM	Crude soluble
LB	BL-21*	37	N/A	12	0.5	0.5 mM	Crude soluble
LB	BL-21*	37	N/A	12	0.5	1.0 mM	Crude soluble

.

One culture at each temperature contained the plasmid in the BL-21 expression host, and the other contained the plasmid in BL-21*. Cultures grown at 37 °C were harvested after 12 h growth while the 30 °C cultures were allowed to grow 16 h before harvesting. The 15 °C cultures were incubated for 30 h and then harvested. Each of the cultures was centrifuged at 4,500 *g* for 15 min to pellet the cells. The cell pellets were stored at -80 °C for protein purification.

STEP 2: CRUDE EXTRACT PREPARATION. All steps were performed at 4 ºC or on ice. To assess gene expression level and protein solubility, crude protein extracts were prepared by sonication and centrifugation. Briefly, cells were resuspended in 1 mL of Lysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole with 10% glycerol pH 8.0). Cells were disrupted by 15 × 2 s sonication pulses with 5 s cooling periods between the pulses. Aliquots of whole cell sample (50 µL) were removed from each sample and retained for later analysis. Crude protein extracts were obtained by centrifugation at 10,000 *g*, at 4 ºC for 20 min.

STEP 3: CRUDE EXTRACT ANALYSIS. Whole cell, pellet, and supernatant samples representing each of the three temperatures and the two types of protein expression cells

were analyzed using SDS-PAGE and Western blot following the procedures specified earlier in this chapter. The Anti-thioredoxin primary and anti-mouse secondary antibodies were used for the Western blotting procedure.

4.2.11 LARGE SCALE IL-12B GENE EXPRESSION IN E. COLI

STEP 1: GROWTH OF E. COLI CELLS. A culture containing 20 mL LB media, supplemented with 100 µg/mL ampicillin was inoculated with Bl-21* cells carrying the pET32a/IL-12b construct. This culture was grown at 37 °C with shaking. After overnight incubation, the culture was used to inoculate 6 L of LB media (3 mL/ L), supplemented with 100 µg/mL ampicillin. Cultures were incubated at 37 °C with shaking overnight. The cells were harvested by centrifugation at 4,500 *g* for 15 min and stored at -80 °C.

STEP 2: CRUDE EXTRACT PREPARATION. All steps were carried out at 4 ºC or on ice. Cells were resuspended in 2 mL Lysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole with 10% glycerol, pH 8.0) per gram cell wet weight. Lysozyme, 1 mg/mL, was added to the cell suspension and incubated for 1 h with stirring every 15 min. Cells were disrupted by 10 × 18 s sonication pulses with 20 s cooling periods between pulses. Crude protein extracts were obtained by centrifugation at 10,000 *g*, at 4 ºC for 20 min.

STEP 3: NI-NTA CHROMATOGRAPHY. Protein was purified by Ni^{2+} affinity chromatography following manufacturer's instructions with 10% glycerol in all buffers. Briefly, 5 mL of resin was prepared by washing with Lysis buffer II and allowed to bind with soluble protein at 4 ºC for 1 h, loaded onto a 2 × 25 cm column. The column was packed by gravity flow and washed with 5 column volumes (CV) of Lysis buffer II followed by two washes with 10 CV of Wash buffer II (50 mM NaH_2PO_4, 300 mM NaCl,

20 mM imidazole, 10% glycerol, pH 8.0). IL-12b protein was eluted with 12 mL of Elution buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 15% glycerol, pH 8.0), which was collected in 1 mL fractions. Whole cell, pellet, flow through, wash, and elution samples were analyzed by SDS-PAGE and Western blotting according to the protocols described earlier in this chapter. Fractions containing IL-12b as determined by SDS-PAGE and Western blot, were collected and dialyzed against 3 L of Dialysis buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 15% glycerol, pH 8.0). The dialyzed protein solution was concentrated using the Amicon filtration system. Protein was flash frozen in liquid nitrogen and stored at -80 ºC until use.

4.2.11.1 In-Gel Tryptic Digestion of IL-12b

The major protein bands were excised from the Coomassie stained gel with a fresh razor blade. The gel slices were processed according to the method of Shevchenko, as specified in section 4.2.3.13.[328] Samples were submitted to the ICMB Protein Mass Spectrometry Facility for MALDI TOF/TOF protein identification.

4.2.12 CLONING OF TMTC2 AND TMTC3 FOR S. CEREVISIAE

4.2.12.1 PCR Reactions and Product Purification

The potential CMT genes *tmtc2* and *tmtc3* genes were amplified using constructs obtained from OPEN Biosystems as template DNA. TMTC2 primers were designed incorporating the restriction enzyme recognition sites for *Eco*RI and *Cla*I in the primer sequences (5′ start primer, 5′-CCAA*GAATTC*AATAATGATTGCAGAGTTGGTGAGC-AGCG-3′, and 3′ halt primer, 5′-GTTGTT*ATCGAT*CCGGTCTTAGAAGTCTTTAAGC-

CTTGTTTTTCC-3′). TMTC3 primers were designed incorporating the restriction enzyme recognition sites for *Apa*I and *Sal*I in the primer sequences (5′ start primer 5′-TAA*GGGCCC*AATATGGCTAATATTAACCTAAAAGAAATAACC-3′ and 3′ halt primer 5′-CCAT*GTCGAC*TTCACCATTTAAAATACGTTCAATC-3′). These primers were designed for the pESC-His *S. cerevisiae* dual expression vector, incorporating Kozak sequences before the translation initiation codons and removing the stop codons from the amplified genes such that a *C*-terminal FLAG- or *myc*-epitope-tag is introduced upon expression, respectively.[332] The primers and template were prepared as previously specified. The PCR reactions consisted of the following: 140 µL of sterile water, 20 µL of 10X KOD buffer, 16 µL of a 2 mM dNTP solution, 12 µL of a 25 mM $MgSO_4$ solution, 10 µL of dimethyl sulfoxide (DMSO), 1.0 µL of each primer (5′ and 3′) at 10 µM, and 1 µL of template DNA, for a total volume of 200 µL. The mixture was divided into two 0.5 mL thin-walled PCR tubes and 2 µL KOD polymerase was added to each reaction. PCR reactions were carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial polymerase activation step of 2 min at 95 ºC. The program consisted of a 20 s denaturing step at 95 ºC, an annealing step at 43.6 ºC for 10 s, and an elongation step at 70 ºC for 90 s. This cycle was repeated thirty times, and the program ended with a 10 min extension at 70 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

4.2.12.2 TMTC2 and TMTC3 S. cerevisiae Construct Shuttle Vector Cloning

PCR products were blunt end ligated into the pCR-Blunt vector following the procedures specified in section 4.2.3.6 with the following specifications. All media was

supplemented with 50 µg/mL kanamycin. LB-agar plates were used for solid culture, and 2X YT media was used for liquid cultures (1.6% tryptone, 1% yeast extract, 0.5% NaCl, pH 7.0). Liquid culture volume was 3 mL and incubation time was 12-16 h. For these genes, five sequencing reactions, using two external primers (M13 forward and reverse) from the cloning vector and three internal primers were submitted to ensure full sequence coverage. Sequences for internal primers used for *tmtc*2 are as follows: GGGAATTCCATATGGATGGGAAACAAACCACC, GGGAATTCCATATGCTCAA-GACTGCGATCAGG, and CCGCTCGAGTTTCTCCAGTCACAAACTGTT. Sequences for internal primers used for *tmtc*3 are as follows: TTGCCTTGACAGTGTTTTTAGTGGC, CCTCCAAGACTGGTTTAATGGCG, and CCTGATCCGAGCAAATGAGTCCCG. Sequences were analyzed as described in previous sections.

4.2.12.3 TMTC2 and TMTC3 S. cerevisiae Construct Expression Vector Cloning

The pESC-His Dual TMTC2, pESC-His Dual TMTC3 and pESC-His Dual TMTC2/3 expression constructs were generated following protocols detailed in section 4.2.3.10. Specific details are as follows. All liquid and solid media was supplemented with 100 µg/mL ampicillin for selection. LB-agar was used for solid cultures and 2X YT media was used for liquid cultures. Restriction enzymes used for digestion were *Eco*RI and *Cla*I for *tmtc*2 and *Apa*I and *Sal*I for *tmtc*3. During CIP treatment of pESC-His Dual vectors, 0.1 µL CIP and 0.25 volumes of 1 M Tris, pH 8 were added to reactions. Ligations were performed at 4 ºC overnight, according to the procedure described in section 4.2.3.10 with the following adjustments to the protocol. For the transformation, 100 µL of competent *E. coli* DH5α was used. Culture volumes for pESC vector

amplification or isolation were 3 mL. Following alkaline lysis, cultures were incubated for 6 h only. Plasmids generated via this method included pESC-His-TMTC2, where TMTC2 is under the control of the pGal10 promoter, *C*-terminally tagged with the FLAG epitope, and pESC-His-TMTC3, where TMTC3 is under the control of the pGal1 promoter, *C*-terminally tagged with the c-*myc* epitope.

To generate pESC-His-TMTC2/3, the latter plasmid pESC-His-TMTC3 was digested with *EcoR*I and *Cla*I and prepared as described above. The digested, CIP treated vector was gel purified and subjected to ligation with *tmtc2*, also digested with these enzymes. Ligation, alkaline lysis, electrophoresis and final plasmid preparation were repeated as above to construct this plasmid. After construction, the pESC-His dual expression vector containing both TMTC2 and TMTC3 was submitted to DNA sequencing to ensure that both genes were present in the vector.

4.2.13 CLONING OF IL-12B WITH SECRETION SEQUENCE FOR S. CEREVISIAE

4.2.13.1 PCR Leader Sequence Design

To target IL-12b for secretion from *S. cerevisiae*, primers were designed to incorporate a signal secretion sequence at the 5′ end of the il-12b gene using TBC methodology.[333] The sequence to be added to the 5′ end of the gene was a three-part sequence beginning with a leader sequence (L), followed by a prepro sequence (P), and ending with a spacer (S) between the prepro sequence and the gene. [334, 335] The protein sequence is: (L) MKLKTVRSAVLSSLFASQVLG (P) QPIDDTESQTTSVNLMADDTESAFATQTNSG GLDVVGLISMAKR (S) EEGEPK. Back translation of the protein sequence yielded the DNA sequence to be added as: 5′-

ATGAAATTGAAAACTGTTAGATCTGCTGTTTTGTCTTCTTTGTTTGCTTCTCAA
GTTTTGGGTCAACCAATAGACGATACTGAATCTCAAACTACTTCTGTTAATTT
GATGGCTGATGATACAGAATCAGCTTTTGCTACTCAGACTAATTCTGGTGGTT
TGGATGTTGTTGGTTTGATTTCTATGGCTAAAAGAGAAGAAGGTGAACCAAA
A-3′.

Primers used to construct this DNA sequence were designed using Helix web analysis found at the website: http://helixweb.nih.gov/dnaworks. The above DNA sequence was entered into the program, along with the first 30 nucleotides of *il-12b* (ATGTGTCACCAGCAGTTGGTCATCTCTTGG) to generate primers for overlap extension PCR, to synthesize the sequence *in situ*. Helixweb analysis generated 8 primers of 60 bases with overlapping regions of approximately 20 nt each, with melting temperatures of 62 ºC. Finally, the restriction site (*Sac*I) and Kozak sequence were added to the 5′ sequence of the ultimate 5′ primer. The 3′ primer, including the *Xho*I restriction site, was designed to match the length, GC content, and melting temperature of the final 5′ primer. The 3′ primer was used to amplify the entire *il-12b* gene in the final PCR step. The final primer sequences are summarized in Table 4.4 below.

4.4 IL-12b *S. cerevisiae* Primers

Primer #	Sequence
5′ / 1	5′-GGAGCTCATTATGAAATTGAAAACTGTTAGATCTGC-3′
2	5′-CAAAACTTGAGAAGCAAACAAAGAAGACAAAACAGCAGATCTAACAGTTTTCAATTTCAT-3′
3	5′-TTCTTTGTTTGCTTCTCAAGTTTTGGGTCAACCAATAGACGATACTGAATCTCAAACTAC-3′
4	5′-CTGATTCTGTATCATCAGCCATCAAATTAACAGAAGTAGTTTGAGATTCAGTATCGTCTA-3′
5	5′-GATGGCTGATGATACAGAATCAGCTTTTGCTACTCAGACTAATTCTGGTGGTTTGGATGT-3′

6	5′-CACCTTCTTCTCTTTTAGCCATAGAAATCAAACCAACAACATCCAAACCACCAGAATTAG-3′
7	5′-CTATGGCTAAAAGAGAAGAAGGTGAACCAAAAATGTGTCACCAGCAGTTGGTCATCTCTT-3′
8	5′-CCAAGAGATGACCAACTGCTGGTGACA-3′
3′	5′-ATACTCGAGCCTCCAAATTTTCATCCTGGATC-3′

4.2.13.2 Overlap Extension PCR Reactions to Generate the IL-12b Leader Sequence

STEP 1: GENERATION OF THE LPS SEQUENCE. The primers were obtained as lyophilized powder and were prepared by dilution with TE buffer to a concentration of 100 µM. This stock solution was further diluted to 10 µM with sterile water. A second dilution was made to 3 µM, and 1.25 µL of each primer 1-8 were combined with 5 µL of 10X buffer, 5 µL of 2.0 mM dNTP, 3 µL of 25 mM $MgSO_4$, 1 µL of KOD polymerase, and sterile water to a volume of 50 µL. The reaction was heated to 95 ºC for 2 min with the lid temperature set to 105 ºC. After the initial denaturing step, the reaction was heated for 15 s at 94 ºC, followed by an annealing step at 57 ºC for 30 s. The temperature was raised to 68 ºC for 15 s and the cycle was repeated 25 times.

STEP 2: AMPLIFICATION OF THE LPS SEQUENCE. The PCR reaction from the last step was used as the template for the next reaction. To this reaction was added: 2.5 µL of 10 mM primers 1 and 8 only, 5 µL 10X buffer, 10 µL of 2.0 mM dNTP, 3 µL of 25 mM $MgSO_4$, 2 µL of KOD polymerase and sterile water to a volume of 100 µL. The PCR cycle described above was repeated, and the reaction was stored at 4 ºC until purification.

STEP 3: GEL PURIFICATION OF THE LPS SEQUENCE. The PCR products were purified by agarose gel electrophoresis as described previously and visualized using a UV illumination source. DNA bands of the correct base pair length (approximately 300 bp) were excised from the gel and purified using the Qiagen gel extraction kit, following the

manufacturer's protocols with the modifications as described previously, except that an elution volume of 20 µL was used.

STEP 4: GENERATION OF THE LPS-IL-12B SEQUENCE. Each primer, (1.25 µL of 5′ and 3′) at 3 µM were combined with 5 µL of 10X buffer, 5 µL of 2.0 mM dNTP, 3 µL of 25 mM $MgSO_4$, 20 µL of the purified DNA from the previous step, 0.1 µL of IL-12b template DNA, 1 µL of KOD polymerase, and sterile water to a volume of 50 µL. The reaction was heated to 95 ºC for 2 min with the lid temperature set to 105 ºC. After the initial denaturing step, the reaction was heated for 15 s at 94 ºC, followed by an annealing step at 57 ºC for 30 s. The temperature was raised to 68 ºC for 1 min and the cycle was repeated 25 times.

STEP 5: AMPLIFICATION OF THE LPS-IL-12B SEQUENCE. The PCR reaction from the last step was gel purified and used as the template for the next reaction. Again, the elution volume of the gel purified product was reduced to 20 µL. The primers, (0.5 µL of 10 mM) 5′ and 3′ only, 2.5 µL of 10X buffer, 10 µL of 2.0 mM dNTP, 1.5 µL of 25 mM $MgSO_4$, 10 µL of DNA template, 0.5 µL of KOD polymerase for a volume of 25 µL. The reaction was heated to 95 ºC for 2 min with the lid temperature set to 105 ºC. The reaction was heated to 95 ºC for 2 min with the lid temperature set to 105 ºC. After the initial denaturing step, the reaction was heated for 15 s at 94 ºC, followed by an annealing step at 57 ºC for 30 s. The temperature was raised to 68 ºC for 1 min and the cycle was repeated 25 times. The cycle terminated in a 10 min final extension at 70 ºC and the reaction was stored at 4 ºC until purification.

STEP 6: GEL PURIFICATION OF THE LPS-IL-12B SEQUENCE. The PCR products were purified by agarose gel electrophoresis employing an 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide as specified in section 4.2.3.4, except that the elution volume was reduced to 20 µL.

PCR products were A-tailed and ligated into the shuttle vector pGEM-T easy for blue-white screening following the procedures described earlier in this chapter. The purified plasmid was sequenced to confirm the identity of the PCR product. For these genes, two sequencing reactions, using two external primers (M13 forward and reverse) from the cloning vector were submitted to ensure full sequence coverage, and DNA sequences were evaluated as described previously.

STEP 7: CLONING OF LPS-IL-12B INTO PYES. Once the DNA sequence was confirmed to be the desired gene, *il-12b*, it was isolated and ligated into the pYES plasmid. The pYES-sc-il12b expression construct was generated following protocols detailed in section 4.2.3.10. Specific details are as follows. All liquid and solid media was supplemented with 100 µg/mL ampicillin for selection. Restriction enzymes used for digestion were *Sac*I and *Xho*I, and a typical reaction mixture consisted of 10 µL of DNA, 5 µL of 10X NEBuffer, 35 µL of sterile water and 1 µL of each enzyme. Ligations were performed at 4 ºC overnight and 100 µL of chemically competent *E. coli* DH5α was used for transformation reactions. Culture volumes for pYES vector amplification or isolation were 10 mL. Cultures and plates were incubated for 8-10 h only. The plasmid generated via this method was pYES-sc-il12b for expression of *il-12b* in *S. cerevisiae* with an *N*-terminal secretion sequence for expression of IL-12b in yeast cells, and secretion of the protein into the media. Positive colonies were amplified and gene-containing plasmids purified for transformation into *S. cerevisiae*.

4.2.14 S. CEREVISIAE TRANSFORMATIONS AND PROTEIN EXPRESSION

To transform the InvSC1 yeast strain, a fresh YPD-agar plate was streaked with glycerol stock prepared as described previously. After overnight incubation at 30 ºC, a

liquid culture of 20 mL YPD media was inoculated using sterile transfer technique. This liquid culture was incubated overnight at 30 ºC with 200 RPM shaking. The OD_{600} was measured and the appropriate amount of the culture was transferred to 50 mL fresh YPD media to obtain a culture with an OD_{600} of 0.4. This culture was allowed to grow for 2-4 h at 30 ºC with 200 RPM shaking. The culture was subsequently transferred into sterile 50 mL conical tubes and centrifuged at 2,500 *g* for 5 min to pellet the yeast cells. The supernatant was decanted and the cells were resuspended in 40 mL of 1X TE buffer (10 mM Tris·HCl, pH 7.5, 1 mM EDTA, pH 8.0). The cells were centrifuged again at 2,500 *g* for 5 min and resuspended in 2 mL 1XLiAc/0.5XTE buffer (100 mM lithium acetate 5 mM Tris·HCl, (pH 7.5), 0.5 mM EDTA, (pH 8.0)). The cell suspension was incubated at room temperature for 10 min and 100 µL aliquots were dispensed into sterile eppendorf tubes for transformation.

The plasmid used to transform the yeast cells (1 µg) was mixed with 100 µg salmon sperm DNA and introduced to the yeast strain. The transformation mix (700 µL) (40% PEG 3350, 60% 1XLiAc/0.5XTE buffer) was introduced to the cells, and mixed well. The transformation mixture was incubated at 30 ºC for 30 min, and 88 µL of dimethyl sulfoxide (DMSO) was added to the mixture. After mixing well, the transformation mixture was heat shocked at 42 ºC for 7 min. Cells were pelleted by centrifugation at 2,500 *g* for 30 s, and the supernatant was removed by aspiration. Transformants were washed with 1X TE buffer and centrifuged again to pellet the cells. Cells were resuspended in 50 µL 1X TE buffer and plated on SC-dropout media. When the pYES cloning vector was used, uracil was omitted from media, while histidine was omitted from yeast transformed with the pESC-His vector. For strains transformed with both vectors, InvSC1 was grown on media without histidine or uracil. Plates were incubated at 30 ºC for 2-3 days until colonies appeared on the plates. Positive

transformants were used to inoculate 10-20 mL of SC dropout media. Cultures were grown overnight at 30 ºC with 200 RPM shaking. After growth, an aliquot of 80% sterile glycerol equal to 25% of the culture volume was introduced into the culture and mixed thoroughly. The fresh glycerol stocks were aliquotted into 1 mL volumes and frozen at -80 ºC until needed.

4.2.15 PROTEIN EXPRESSION IN S. CEREVISIAE

Protein expression was carried out following the manufacturer's suggested protocols. Briefly, 200 µL of a glycerol stock was used to inoculate a 20 mL culture of SC dropout media, which was then incubated at 30 ºC with 200 RPM shaking overnight. SC-his was used for TMTC2, TMTC3 and TMTC2/3 expression. SC-ura media was used for IL-12b expression, and for strains containing both plasmids, double dropout SC-ura-his media was used for culturing. After overnight incubation, the OD_{600} of each culture was taken, and an aliquot of each culture was transferred into a sterile tube such that when resuspended in fresh induction media the OD_{600} of the dilution would be 0.4. The aliquots were centrifuged at 2,500 *g* for 5 min to pellet the cells. The media was discarded by decanting and the pellets were washed with 1X TE buffer or 100 mM potassium phosphate buffer, pH 7.4. Cells were repelleted by centrifugation as above and the wash discarded. SC induction media (0.67% yeast nitrogen base, 2% galactose, 0.075% complete supplemental media amino acid mixture, without uracil and histidine, 0.05% uracil, and/or 0.1% histidine) was used to resuspend the cells to an OD_{600} of 0.4. Cells were grown for 14 h, with $1/7^{th}$ culture volume taken as aliquots at 0, 4, 6, 8, 10, 12 and 14 h after induction. These aliquots were centrifuged to pellet the cells at 2,500 *g* for

5 min. The supernatants were discarded and cell pellets stored at -80 ºC until screened for the presence of target proteins.

4.2.16 PROTEIN EXTRACTION FROM S. CEREVISIAE

The cell pellets obtained in the previous step were processed using several methods to determine the optimal extraction process for these proteins. The first methods tried were the total protein extraction protocol (Yaffe-Schatz), the membrane protein extraction protocol (Schutzbach), and total protein extraction using the Y-Per protein extraction reagent.[331, 336]

METHOD I: YAFFE-SCHATZ TOTAL PROTEIN EXTRACTION. The total protein extraction protocol (Yaffe-Schatz) was applied to 1 mL of yeast cells grown in SC induction media. The cells were chilled for 10 min on ice, followed by the addition of 150 µL YS-lysis solution (2M NaOH, 8% β-mercaptoethanol (v/v)). After addition, the cells were mixed well by inversion and incubated for an additional 10 min on ice. trichloroacetic acid (150 µL of a 50% solution) was added and the cells were again incubated on ice for 10 min. The lysed cells were pelleted by centrifugation for 2 min at maximum speed (16,100 *g*) in a tabletop centrifuge at 4 ºC. The supernatant was aspirated and the protein-containing pellet was washed with ice-cold acetone. After an additional 2 min centrifugation step, the wash was removed by aspiration and the pellet was resuspended in 100 µL 2X SDS sample buffer. In pellets where the sample buffer turned yellow, indicating the pH of the solution was too acidic, 10 µL of 1 M Tris, pH 8.0 was added and mixed well. An aliquot (10 µL) of this sample was loaded onto an SDS-PAGE for separation and assessment of protein expression through Coomassie staining and Western blotting.

METHOD II: MEMBRANE PROTEIN EXTRACTION. The membrane protein extraction protocol was adapted from Schutzbach et al. (1993).[331] Cell pellets were washed with 3/10 culture volume water, followed by a 1/10 volume wash with buffer B (25 mM sodium phosphate, pH 8.0, 5 mM $MgCl_2$ and 0.2% β-mercaptoethanol (v/v)). Cell pellets were frozen at -80 ºC overnight and resuspended the following day in 1/100 culture volume of Buffer B. The suspension was sonicated to break the cells for 1 min per 10 mL culture, with 10 s pulses followed by 30 s cooling intervals. Sonication does not always break yeast cells, thus, homogenization using a Potter-Elvejhem-type homogenizer equipped with a tight-fitting (low-clearance) pestle was also used with 20-30 strokes per aliquot. Cell breakage was assessed either by microscopy, or measuring the OD_{600} of the homogenized culture. When the OD_{600} was reduced to less than 20% of the original optical density, the culture was taken to be fully homogenized. The disrupted suspension was centrifuged at 39,000 *g* for 20 min at 4 ºC to obtain a pellet enriched with membrane proteins. The pellet was washed twice with 1/40 volumes of Buffer B. The pellet was resuspended in 1/200 volumes of Buffer B, which was diluted in 1/60 volumes of Buffer C (10 mM sodium phosphate, pH 8.0, 0.5 mM EDTA, 10% glycerol, 0.1% Nonidet P-40, and 1 mM dithiothreitol), and further diluted with 1/400 culture volumes of 10% P-40. The mixture was again centrifuged at 39,000 *g* for 20 min at 4 ºC to obtain a pellet. The supernatant was retained and the pellet solubilized in 1/50 culture volume of Buffer D (100 mM Tris-acetate, pH 7.5, 10% glycerol, 0.2% β-mercaptoethanol, and 0.1% sodium dodecyl sulfate (SDS)). This mixture was again sonicated for 1 min per 100 mL culture volume to solubilize the pellet more thoroughly, and 1/400 culture volume of 10% Nonidet P-40 was added. The mixture was centrifuged again as above, and the supernatant was retained. The process of resuspension in Buffer D, sonication, and addition of Nonidet P-40, followed by centrifugation was repeated and solubilized

fractions combined and 10 µL of these fractions were loaded onto an SDS-PAGE for assessment of protein production by Coomassie staining and Western blot analysis.

METHOD III: TOTAL PROTEIN EXTRACTION EMPLOYING Y-PER. Y-Per reagent from Pierce was applied to *S. cerevisiae* cultures according to the manufacturer's protocols. Briefly, cells were pelleted at 3,000 g for 5 min at 4 ºC. Cells were frozen at -80 ºC until ready to prepare. At this time, cells were weighed and the appropriate amount of Y-Per reagent was added according to the wet weight of the cells, (100 mg cells = 250-500 µL). The mixture was triturated until homogenous and incubated for 20 min at room temperature. After incubation, cell debris was pelleted at 14,000 *g* for 20 min at 4 ºC. The supernatant was removed and analyzed by SDS-PAGE and Western blotting.

4.2.17 SMALL-SCALE MICROSOMAL PREPARATION

Small scale microsomal preparations, purified following the method of the Aebi lab, were used to assess TMTC2, TMTC3 and TMTC2/3 production in yeast.[337] *S. cerevisiae* cells were grown as stated in section 4.2.3.2. Cells were centrifuged at 5,000 *g* for 5 min at 4 ºC and pellets were washed with 100 mM potassium phosphate buffer (pH 7.4). Cells were resuspended in spheroplasting buffer 2 (1 M sorbitol, 5 mM $MgCl_2$, 10 mM DTT, 50 mM HEPES pH 7.4) to a calculated cell density of 200 OD_{600} per mL. To this was added 50 µL of Zymolyase enzyme (75 units) and the solution was incubated at 30 ºC with 200 RPM shaking for 1-2 h. The solutions were centrifuged at 5,000 *g* for 10 min at 4 ºC. The supernatant was removed by aspiration and spheroplasts were lysed by the addition of 20 volumes of Spheroplast lysis buffer (10 mM HEPES, pH 7.4, 2 mM $MgCl_2$, 1 mM PMSF) and incubated on ice for 30 min. After incubation, spheroplasts were homogenized using 12-25 strokes of a Potter-Elvejhem homogenizer. Sucrose

solution (2 M) was added to the homogenate to bring the final sucrose concentration to 0.25 M. The homogenate was mixed well and centrifuged at 2,500 *g* for 10 min to remove cell debris and any unbroken spheroplasts. The supernatant was subjected to ultracentrifugation at 150,000 *g* (avg) for 20 min at 4 ºC, using a Beckman TL-100 centrifuge equipped with a Ti-100.4 rotor. Microsomes were resuspended in 50 mM Tris•HCl pH 7.4, flash frozen and stored at -80 ºC until analysis.

4.2.18 LARGE-SCALE PREPARATION OF YEAST MICROSOMES

4.2.18.1 Preparation of Spheroplasts

Several protocols were used to extract endoplasmic reticulum as microsomal fractions from *S. cerevisiae*. The spheroplasting protocol employed was based on the procedure reported by Shimoni and Scheikman.[338] *S. cerevisiae* cells expressing TMTC2, TMTC3 or TMTC2 and 3 were grown as reported in section 4.2.3.2. These cells were harvested at the appropriate time points for maximum gene expression (as assessed by previous time course assays). After growth, cells were transferred to 50 mL conical tubes and washed with 100 mM potassium phosphate buffer, pH 7.5. The cells were incubated on ice for 10 min and pelleted again at 4,000 *g* for 5 min. The supernatant was decanted and the cells were resuspended to 100 OD_{600} units per mL in MP buffer (100 mM Tris·HCl, pH 9.4, 10 mM DTT) and incubated at room temperature for 10 min. Cells were centrifuged to re-pellet at 18,600 *g* for 30 s and resuspended to 100 OD_{600} units per mL in lyticase buffer (0.7 M sorbitol, 10 mM Tris·HCl (pH 7.4), 1 mM DTT, and 20 mM sodium azide). At this point, a small aliquot (10 µL) was removed and retained separately for subsequent OD_{600} comparison. Zymolyase enzyme was added to the cells at ~15 units/OD_{600} cells. The samples were incubated for approximately 100 min at 30 ºC with

200 RPM shaking. After this time, the OD_{600} was measured to be 26% of the original value, as compared to the untreated aliquot. Samples were chilled on ice for 2 min and centrifuged at 5,200 *g* for 10 min. The supernatant was carefully removed and the pellet was resuspended in 2X JR lysis buffer (0.4 M sorbitol, 100 mM potassium acetate, 40 mM HEPES buffer (pH 7.4), 4 mM EDTA (pH 8.0)) to an OD_{600} of 250 units/mL. The samples were centrifuged at 18,600 *g* for 5 min at 4 ºC. The supernatant was removed by aspiration, and the pellet was resuspended in 2X JR lysis buffer to an OD_{600} of ~1000 units/mL, and samples were flash frozen in liquid N_2 and stored overnight at -80 ºC.

4.2.18.2 Microsomal preparation from spheroplasts

4.2.18.3 METHOD I

Microsomes were prepared using the spheroplasts prepared in the previous step. Following the method of Sanyal et al., the remaining steps were carried out quickly and on ice, or if possible at 4 ºC.[339] Spheroplasts were thawed on ice and an equal volume of sterile water was added to the sample, with 1 mM each of DTT and phenylmethylsulfonyl fluoride (PMSF). Spheroplasts were disrupted with 10-20 strokes of a hand-held Dounce or Potter-Elvehjem homogenizer. Breakage was examined using a microscope. If upon observation, spheroplasts were not broken, homogenization was repeated until broken spheroplasts constituted a majority in the sample. Broken cell suspensions were centrifuged at 1,900 *g* at 4 ºC for 10 min. The supernatant was further clarified by centrifugation at 15,100 *g* at 4 ºC for 10 min. The supernatant was further clarified by centrifugation at 48,500 *g* for 30 min at 4 ºC to pellet microsomes. The pellet was resuspended gently in microsome buffer (20 mM HEPES-NaOH, pH 7.4, 100 mM NaCl).

The microsomal samples were aliquoted, flash frozen in liquid N_2 and stored at -80 ºC until use.

4.2.18.4 METHOD II

Following the method of Shimoni and Schenkman,[338] using the previously prepared spheroplasts, the remaining steps were carried out quickly at 4 ºC. Lysed spheroplasts were thawed on ice and an equal volume of sterile water was added to the sample, with 1 mM each of DTT and phenylmethylsulfonyl fluoride (PMSF). Spheroplasts were disrupted with 10-20 strokes of a hand-held Dounce or Potter-Elvehjem homogenizer. Cell breakage was examined using a microscope. If cells were not broken upon observation, homogenization was repeated until the majority of cells were observed to be broken. Broken cell suspensions were centrifuged at 1,900 *g* for 10 min at 4 ºC to pellet cell debris. The supernatant was retained and centrifuged at 27,200 *g* at 4 ºC. The pellet was resuspended gently in buffer B88 (20 mM HEPES, pH 6.8, 250 mM sorbitol, 150 mM potassium acetate, 5 mM magnesium acetate) to a calculated OD_{600} of 2500 units/mL using a Potter-Elvejhem homogenizer or trituration with a micropipette. The suspension was layered on top of a 1.2/1.5 M sucrose 2 step gradient in 4 mL thick-walled polycarbonate centrifuge tubes. To establish this gradient, 1 mL of a 1.5 M sucrose solution is dispensed into the centrifuge tube. To establish the 1.2 M step, 1 mL of the solution was gently layered on the top of the 1.5 mL solution by running the solution slowly and carefully down the side of the centrifuge tube. Subsequently, 500 to 750 µL of the microsomal sample was added to the sucrose gradient following the methodology used to add the 1.2 M layer. These tubes were carefully balanced to within 0.01 g and centrifuged at 150,000 *g* (avg) at 4 ºC for 1 h. After centrifugation, the 1.2 M layer and remaining solution was aspirated from the top of the tube, taking care not to

disturb the microsomal band located at the 1.2-1.5 M interface. This layer was removed with a Pasteur pipette and diluted with 4-5 volumes of B88 solution. The resulting mixture was centrifuged at 27,200 *g* at 4 ºC for 10 min. The supernatant, which contains sucrose from the ultracentrifugation step was removed and the pellet resuspended in a small volume of B88 buffer. The volume used was approximately equal to the volume of the resulting pellet in the previous step. This pellet was homogenized using a small Potter-Elvejhem homogenizer. The OD_{280} of a 1:100 dilution of the samples with 2% SDS (w/v) were measured. The samples were subsequently diluted to an approximate of 40 calculated OD_{280}, which corresponds to ~ 8 mg/mL total protein. The microsomal samples were aliquotted, flash frozen in liquid nitrogen, and stored at -80 ºC until use.

4.2.18.5 S. cerevisiae Microsomal In Vitro C-Mannosylation Assays

Reactions were carried out using the method of Doucey.[315] The peptide N-AC-KPPQFAWAQWFE-NH_2, as designed based on the RNase2 sequence, replacing Thr-6 with Ala to prevent possible *O*-glycosylation, and used as the mannosyl acceptor substrate.[126] The peptide was incubated in the presence of microsomes from rat liver, yeast and insect cells, followed by HPLC purification using reversed phase HPLC equipped with a C-18 column (acetonitrile/TFA gradient).

The reaction mixture contained in a final volume of 240 µL: 0.9 mM of peptide (~ 400 µg), 145 µg of *S. cerevisiae* microsomal protein, 20 mM HEPES-NaOH, pH 7.2, 110 mM potassium acetate, 2 mM magnesium acetate, protease inhibitor cocktail (leupeptin, 2.4 µg/mL; antipain, 4 µg/mL; lima bean trypsin inhibitor, 40 µg/mL; benzamidine, 20 µg/mL; AEBSF HCl, 10 µg/mL; aprotinin, 2 µg/mL; chymotrypsin, 10 µg/mL; pepstatin A, 5 µg/mL; 2 mM EDTA) and 0.2% v/v Triton X-100. The Triton X-100/protein ratio

(wt/wt) was kept constant at 0.34 (total Triton X-100 concentration, 3 mM). The mixture was incubated at 37 °C overnight.

After overnight incubation, the samples were filtered through a YM-10 centrifugal filter at 14,000 *g* for 20 min at room temperature. The peptide was purified from each sample using HPLC, with a reverse phase C-18 column. Buffer A was water with 0.1 % TFA, and buffer B was 90% acetonitrile, with 0.1 % TFA. The gradient was linear, increasing buffer B at 1%/min, and samples were monitored at 280 nm (tryptophan abs). Peaks were collected and lyophilized to concentrate fractions. The concentrated fractions were submitted for MALDI-MS analysis.

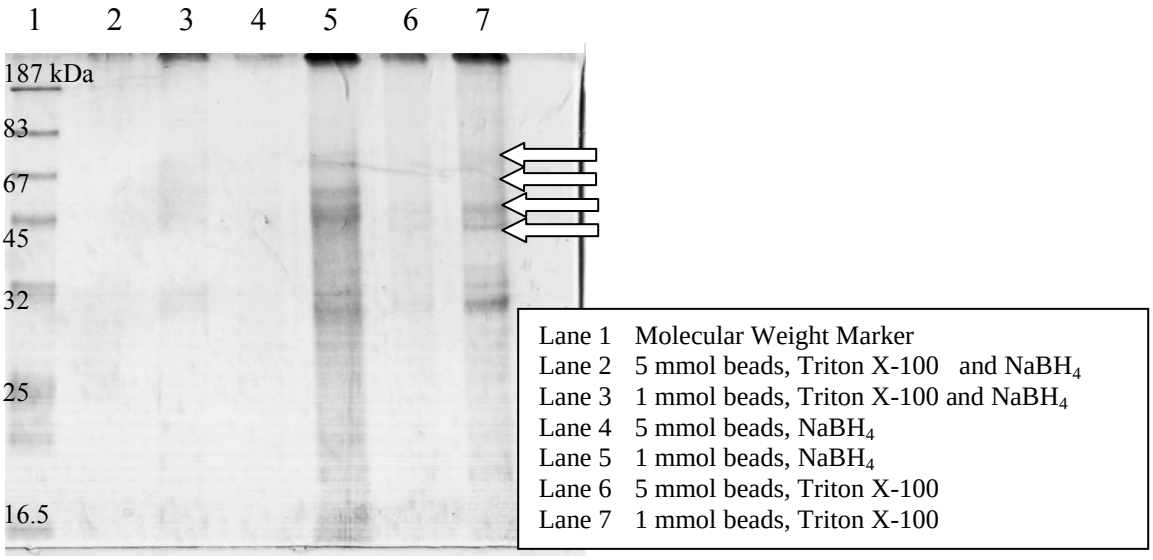
4.3 RESULTS AND DISCUSSION

To isolate the *C*-mannosyltransferase, attempts were made to crosslink a substrate peptide to the enzyme. Purified microsomes from rat liver were allowed to react with the substrate peptide containing a selectively oxidizable 5-hydroxytryptophan residue in the presence of sodium periodate, which is expected to oxidize 5-hydroxytryptophan such that the 5-keto group of 5-HT can crosslink to any reactive amino acid such as the α-amino or ε-amino of Lys, imidazole of His, or thiol of Cys in the enzyme active site.[340-342] The linkage could then be chemically reduced in the presence of sodium borohydride and form an irreversible covalent linkage. The bead-linked peptide could be centrifuged to collect the proteins cross-linked to the peptide and washed to remove any unlinked proteins. The peptide was released from the agarose beads by boiling and the proteins crosslinked to the peptides could be separated on 12% denaturing acrylamide gel.

Strong bands were excised from the gel, trypsin digested (in-gel) and submitted for mass spectrometry analysis. A significant hit from the MS results was the *R.*

norvegicus (rat) dolichyl-diphosphooligosaccharide-protein glycosyltransferase (P25235) 63 kDa subunit precursor (EC 2.4.1.119). While this was not the protein we had expected to isolate, we began to examine what was known about *C*-mannosylation and whether this information would assist us in identifying the protein *in silico*. Previous research has shown that the *C*-mannosyltransferase protein is targeted to the endoplasmic reticulum and thus has a distinct ER signal sequence.[315] In addition, the CMT uses dolichol phosphate mannose as the substrate.[315]

4.1 CMT Crosslinking Gel



Glycosyltransferases using dolichol-phosphate as a substrate have a multispan transmembrane topology.[343] Interestingly, the family of dolichol-phosphate mannose protein *O*-mannosyltransferases (POMT), is highly conserved from yeast to humans.[91] Studies of the POMT structure and function have identified two domains critical for

mannosyltransferase activity.[225] The first domain, the protein mannosyltransferase (PMT) domain, is found in all dolichol-phosphate mannose protein transferases and is thought to be the binding site for the glycolipid. [225] The second domain is the mannosyltransferase, inositol and ryanodine receptor (MIR) domain, which has been postulated to be the ligand binding/transfer domains found in the three proteins for which they are named.[344] Human POMTs have a defined structure consisting of 9 transmembrane domains. [225, 345] The PMT domain and three (MIR) domains are located on hydrophilic loops between the 1st and 2nd and 8th and 9th transmembrane domains, respectively. These domains are believed to come together to catalyze the mannosyltransferase reaction. [225]

Sequence alignment cannot distinguish between *C*-glycosyltransferases and *O*-glycosyltransferases in prokaryotes. A recent paper demonstrated that the prokaryotic urdamycin *C*-glycosyltransferase (UrdGT2) displays the unusual ability to generate both C–C and C–O glycosidic bonds. [84] It is possible that this sequence similarity may carry over into eukaryotic glycosyltransferases. The *C*-mannosyltransferase (CMT), since it has a similar function to the POMT, should show some sequence similarity to POMTs. To identify the human CMT gene or genes, we searched for "hypothetical" or "predicted" proteins in the human genome that contained the conserved PMT domain.

H. sapiens protein *O*-mannosyltransferase 1 POMT-1 (gi 34190043) was subjected to pblast using the NCBI "Blast" website. Using all possible protein builds in the database, 23 proteins were identified as having some homology to POMT-1. Of the proteins identified, the majority (22) of these proteins were identified as POMT-1 splice variants, POMT-2, or fragments of either of these two proteins by manual sequence alignment using the NCBI website protein alignment tool. Discarding all splice variants or fragments of POMT1 or POMT2 left one protein, annotated TMTC2 (transmembrane

and tetratricopeptide repeat containing #2), which has a non-canonical PMT domain and does not correspond to the known POMT domain architecture (PMT, MIR domains).

This protein was present in four builds used for protein analysis and was found under the following gi numbers: gi|62739786|gb|AAH93852.1| hypothetical protein LOC160335, gi|22749211|ref|NP_689801.1| hypothetical protein LOC160335 [Homo sapiens], gi|62740007|gb|AAH93854.1| hypothetical protein LOC160335 [Homo sapiens], and gi|21740314|emb|CAD39165.1| hypothetical protein [Homo sapiens]. As shown in figure 4.2, the TMTC2 protein contains a non-canonical PMT domain.

4.2 TMTC2 PMT Alignment with Canonical PMT Sequence

```
          10        20        30        40        50        60
   ....*....|....*....|....*....|....*....|....*....|....*....|
43 WTHIFYND---FWGTLLTHSGSHKSYRP--LCTLSFRLNHAIGGLNPWSYHLVNVLLhAA
36 ANALFFMDvhpPLGKMLIALGGYLAGKPplFYFISSGSFLYFGNVPYFSMRLFSALL-GS
          80        90        100       110       120       130
   ....*....|....*....|....*....|....*....|....*....|....*....|
97 VTGLFTSFSKILLGDGYWT-FMAGLMFASHPIH-TEAVAGIVgraDVGASLFFLLSLLCY
95 LTVPLVYLLLLKLGFSRLAaLLAALLVALDNSFvTLSRYILL---DSPLLFFTTLAMYCL
          160       170       180       190       200
   ....*....|....*....|....*....|....*....|....*....|
55 IKHCSTRGySARTWGWFLGSGLCAGCSMLWKeqGVTVLAVSAVYDVFVFH 205
52 LKFEKAPF-SRKWWLWLLLTGIALGLAVSTK--GVGLFTVLPVGLLFIWL 198
```

This protein is also lacking the MIR domains found in POMTs. The *C*-terminus of the protein contains multiple TPR (tetratricopeptide) repeats. The TPR is a coiled coil structure that creates an amphipathic channel, which is believed to function as a protein-protein interaction domain that binds the α-helices of multiple proteins. [208] This is noteworthy because the protein responsible for mono β-*O*-GlycNAcylation (β-OGT) of many cytosolic proteins also has a TPR binding motif. [207] Clearly, this protein which exhibits low homology to POMT-1 is a promising candidate worthy of further investigation.

Indeed, further analysis of this protein revealed that β-OGT is a close paralog of TMTC2. In other words, TMTC2 has homology to POMTs and also shows high similarity to but a function different than β-OGT. Thus, the TMTC2 protein is likely to have glycosyltransferase activity, although TMTC2 most likely has a divergent function from the other two proteins. Upon closer examination, the *N*-terminal portion of the PMT domain of TMTC2 is truncated. Interestingly, this region has been determined to be important for protein binding in other PMT domain-containing proteins. POMT1 and 2 are thought to complex with each other in the ER using the *N*-terminal region of the protein.[225] It has been proposed that the POMT1/2 complex flips Dol-P-man from the cytoplasmic face of the ER to the lumen so that the mannose residue is accessible to the MIR catalytic domains.[225]

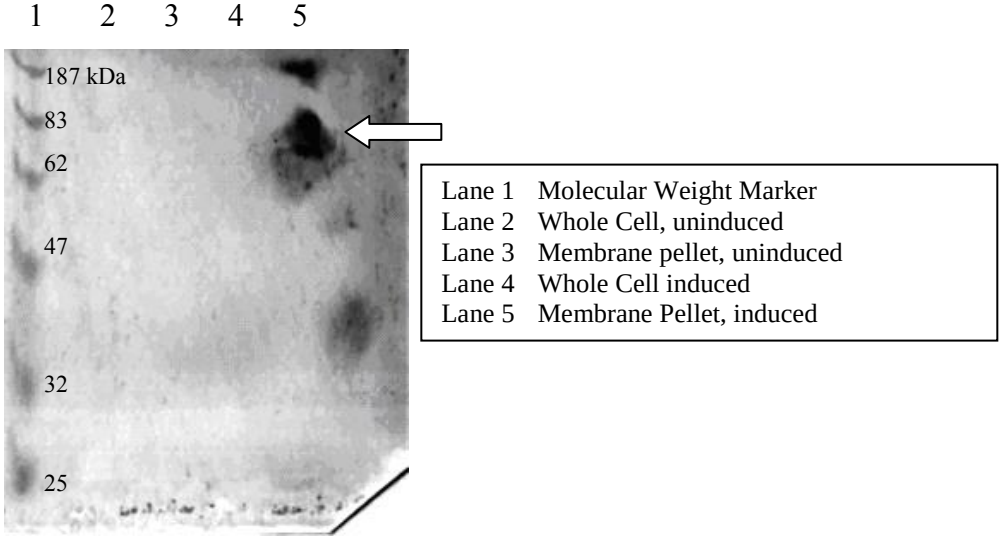
Investigating the TMTC2 protein using other databases (Uniprot, DBGET, Interpro, Pfam, Prosite), revealed several important factors. Using the LIFEdb database, the subcellular location of the protein was confirmed to be the endoplasmic reticulum.[346] Protein analysis *in silico* also afforded the identification of another domain in place of the three MIR domains. This newly identified domain, DUF1736, is a "domain of unknown function" that spans about 100 amino acids and is in the location of the proposed catalytic domain in POMT1/2. The secondary structure of TMTC2 was analyzed and compared to the known structure of the yeast POMT protein using the PHYRE structure prediction server. With very few exceptions, the secondary structure for the entire protein is predicted to be α-helical (no β-sheets) with a large unstructured hydrophilic loop (luminal) in the middle of the DUF domain that corresponds to the positioning and proposed random coil structure of the active site in the POMT proteins. Membrane topology predicts this loop to be in the lumen of the ER.

NCBI CDART, the conserved domain architecture recognition tool, was then used and 14 proteins were identified with structure identical to TMTC2 protein architecture. The results indicated that TMTC2 homologs exist in: *Rattus norvegicus, Mus musculus, Macaque macaca fascicularis, Canis familiaris, Xenopus laevis, Bos taurus, Gallus gallus, Danio rerio, Tetraodon nigroviridans, Caenorhabditis elegans, Caenorhabditis briggsae, Drosophila melanogaster, Drosophila pseudoobscura, Apis mellifera, Anopheles gambiae, Tribolium castaneum, Strongylocentrotus purpuratus, Blastopirellula marina, Trichodesmium sp.,* and *Synechocystis sp.*. Notably, no fungal, bacterial strains or plant species have this gene. In contrast, higher order eukaryotes; mammalian, avian, amphibian and annelid model organisms all contain this gene. Using the Uniprot server to align protein homologs against TMTC2, it was determined that the proteins in the marine organisms do not contain a canonical DUF 1736 domain, although it has been reported that some marine organisms produce *C*-mannosylated natural products.[316] In contrast, the TMTC2 homologs in insects (honeybee, *Drosophila* and *Anopheles*) contain the DUF 1736 domain and TPR repeats but do not have the canonical PMT domain. It is possible that the PMT domain in insects is different than that of the canonical PMT sequence for higher organisms. It should be noted here that the *Drosophila* rotated abdomen protein, which is a predicted POMT, does not have an apparent PMT domain.

Overall, the presence of the PMT domain in TMTC2 suggests that TMTC2 most likely uses Dol-P-Man as the substrate and it has the ability to bind to many different protein substrates due to the presence of the TPR region, which contains 10 TPR repeats. On the basis of these analyses, we decided to carry out more experiments to investigate TMTC2 as a potential *C*-mannosyltransferase.

The gene TMTC2 was obtained from OPEN Biosystems and used to create constructs for expression in *E. coli*. The gene *tmtc2* was expressed in *E. coli* under 56 different conditions. Similar work was performed by the Imperiali group, who expressed the *alg2* and *alg11* genes as thioredoxin fusion transmembrane proteins in *E. coli*.[109] Membrane fractions were difficult to load in acrylamide gels, even after solubilization with 8 M urea or upon the addition of 10% Nonidet. Samples often ran as 'smears' in the gels as indicated by Coomassie staining. No distinct band corresponding to the desired molecular weight of TMTC2 was observed in these gels. However, in the Western blot, positive signals appeared as faint smears, which were observed under multiple induction conditions. Thus, there was a low level of *tmtc2* expression. The low level expression is likely due to the observation of premature cell death. This is not surprising as work done by Bruno and Miroux suggested that expression of transmembrane proteins in *E. coli* caused cell death.[347]

4.3 TMTC2 Production in *E. coli*



4.3.2 IL-12B EXPRESSION IN E. COLI

Results from the previous small-scale experiments demonstrated that the best expression and solubility conditions were 37 °C in Bl-21* cells with 1 mM IPTG induction, although our previous Western Blot indicated that IL-12b protein had low expression under these conditions. Large-scale preparations using these conditions yielded IL-12b as indicated by Western blot. Protein concentration was not determined as the protein was only 50% pure using this procedure. IL-12b has a molar extinction coefficient (ε) of 66970 $M^{-1}cm^{-1}$ at A_{280} and a molecular weight of 37.3 kDa. The Il-12b protein purified by this procedure was of low quality and could not be used in *in vitro* assays.

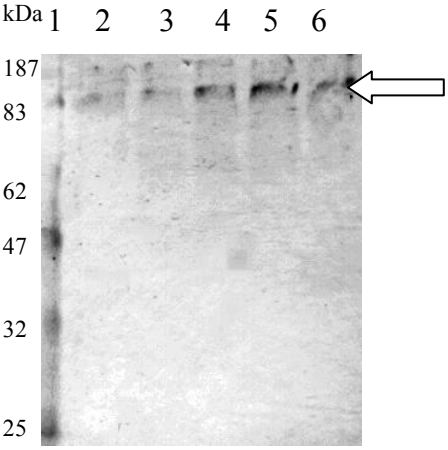
4.3.3 IL-12b Expression in S. cerevisiae

Attempts were then made to express the protein in *S. cerevisiae*. Yeast is an excellent platform for *C*-mannosylation work, as it is a eukaryotic system and thus naturally produces the substrate dolichol phosphate mannose (Dol-P-man), and there is no endogenous *C*-mannosyltransferase.[228] The expression level of *tmtc2* and *tmtc3* in yeast cells was much higher. Our experimental design entailed the coexpression of the TMTC2 proteins (TMTC2, TMTC3, and TMTC2/3) with IL-12b, a protein known to be naturally *C*-mannosylated *in vivo*.[348] Modified IL-12b should be secreted into the yeast culture media, where it can be harvested and analyzed for protein modifications. Unfortunately, *S. cerevisiae* does not appear to secrete IL-12b well. Protein targeting is also a concern in this system. If the ER targeting signal intrinsic to the TMTC2 and TMTC3 proteins is not recognized by the *S. cerevisiae* translocation machinery and correctly translocated into the endoplasmic reticulum, then the proteins will not have access to their native substrate (Dol-P-man) and will not demonstrate activity. Due to

these difficulties, we chose to express *tmtc2*, *tmtc3* and *tmtc2/3* in yeast strains and purify the ER fractions as microsomes for use in *in vitro* assays.

Purification of all three proteins (TMTC2, TMTC3 and IL-12b) was attempted and it was clear that these proteins did not express well in *S. cerevisiae*. However, we were able to achieve expression of *tmtc2* and *tmtc3* in yeast, as shown in the Western blots below in Figures 4.4 and 4.5, respectively.

4.4 TMTC2 Production in *S. cerevisiae* (Time Course Assay)

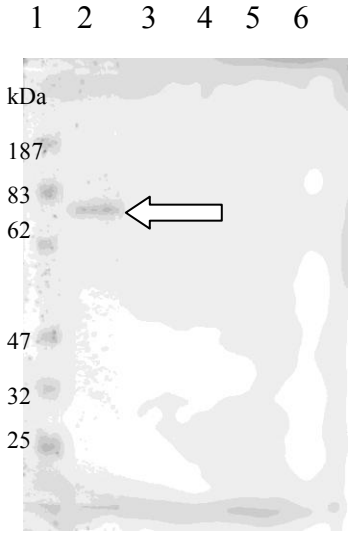


Lane 1	Molecular Weight Marker
Lane 2	0 h growth, soluble YS prep
Lane 3	2 h growth, soluble YS prep
Lane 4	4 h growth, soluble YS prep
Lane 5	6 h growth, soluble YS prep
Lane 6	8 h growth, soluble YS prep

Cell samples were taken at 0h, 2h, 4h, 6h, 8h and 10h and processed according to the Yaffe-Schatz method for total protein isolation. The best expression results for TMTC2 were achieved 8-10 h post-induction in *S. cerevisiae*.

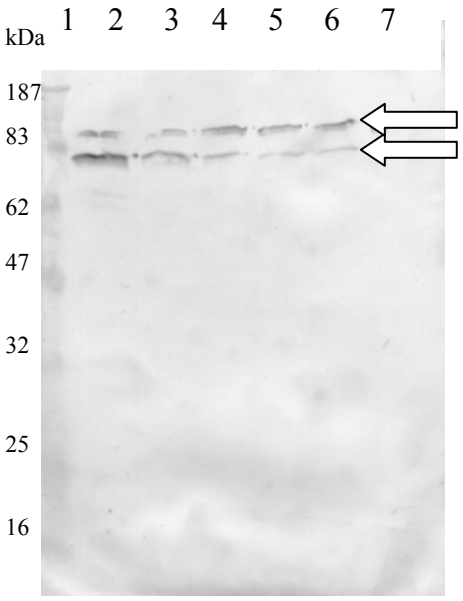
Samples were taken at 0h, 2h, 4h, 6h, 8h and 10h and the total protein was extracted using the Yaffe-Schatz method. For TMTC3, the best protein production was achieved 10 h post-induction in *S. cerevisiae*, as shown below.

4.5 TMTC3 Production in *S. cerevisiae* (Time Course Assay)



Lane 1	Molecular Weight Marker
Lane 2	10 h growth, soluble YS prep
Lane 3	8 h growth, soluble YS prep
Lane 4	6 h growth, soluble YS prep
Lane 5	4 h growth, soluble YS prep
Lane 6	0 h growth, soluble YS prep

4.6 TMTC2 + IL-12b Production in *S. cerevisiae*



Lane 1	Molecular Weight Marker
Lane 2	10 h growth, soluble YS prep
Lane 3	8 h growth, soluble YS prep
Lane 4	6 h growth, soluble YS prep
Lane 5	4 h growth, soluble YS prep
Lane 6	2 h growth, soluble YS prep

Cell samples were taken at 0h, 2h, 4h, 6h, 8h and 10h post induction. The best expression results for the TMTC2 protein under these conditions were achieved 10 h post-induction in *S. cerevisiae*. IL-12b could not be seen on the Western blot, thus we abandoned our *in vivo* assay design, and turned instead to designing an assay to test these enzymes *in vitro*.

Hofsteenge established an *in vitro C*-mannosylation system using microsomes as the source of enzyme on the basis of the observation *in vivo*, that the first 12 residues of RNase 2 are sufficient for modification.[314] Reactions were carried out after the method of Doucey.[315] The peptide substrates, N-KPPQFAWAQWFE-C were reacted with microsomal fractions from *S. cerevisiae* containing TMTC2 only, TMTC3 only, TMTC2/3, or no protein. The reaction mixture contained reaction buffer (20 mM HEPES-NaOH pH 7.2, 110 mM potassium acetate, 2 mM magnesium acetate), protease inhibitor cocktail, 2 mM EDTA and 0.2% Triton X-100. The Triton X-100 to protein ratio (wt/wt) was kept constant at 0.34 (total Triton X-100 concentration 3 mM). The mixtures were incubated at 37 °C overnight. Peptides were removed from the reaction by size exclusion filtration and HPLC equipped with a reverse phase C-18 column, using a linear gradient of water/acetonitrile (0.1%) TFA. HPLC Peaks with absorbance at 214 and 280 were collected and concentrated. These samples were submitted for MS analysis.

Mass spectrometry results were negative as no *C*-mannosylation was found. These peptide samples submitted were still mixtures, despite HPLC purification, and thus it was not possible to identify the peptide substrate in control samples.

4.4 CONCLUSIONS

In this chapter, we identified proteins TMTC2 and TMTC3 as putative *C*-mannosyltransferases. However, we were unable to demonstrate the activity of the proteins TMTC2 and TMTC3 in bacterial platforms. Bacterial expression of these genes failed under more than 50 conditions, which differed in media composition, temperature, cell lines and induction conditions. Transmembrane proteins are difficult to produce in *E. coli* and other bacteria as the overproduction of these proteins is theorized to overload the translocation system that is so essential for cellular metabolism.[347, 349] Thus, attempts were made to produce the TMTC2 and TMTC3 proteins in *S. cerevisiae*.

Expression of the genes *tmtc2* and *tmtc3* were achieved in *S. cerevisiae*. However, co-expression of the target gene *il-12b* was unsuccessful, which nullified the original assay design. *In vitro* assays using yeast microsomes were also carried out. Synthetic peptides were made and tested with microsomes containing TMTC2 and TMTC3. MS analysis failed to show *C*-mannosylation of the peptides. We examined the myriad reasons for these results. It is quite possible that the proteins in question. TMTC2 and TMTC3, are not targeted to the ER in yeast, and thus are not present in the microsomes harvested from the cells. This would also invalidate any *in vivo* assays we could perform in *S. cerevisiae*. It is also possible that protein misfolding has occurred, and even if directed to the proper intracellular organelle, the protein(s) may be non-functional. The *O*-GlcNAc transferase β-OGT has a tetratricopeptide repeat region and has been shown to be active when expressed in insect cells.[209] *ß-ogt* expression has never been reported in yeast or bacteria. Is it possible that bacteria or yeast do not have the capacity to properly fold a massive tetratricopeptide repeat region, as found in β-OGT, TMTC2 and TMTC3?

It is also possible that the protein is not functional under cellular conditions in *S. cerevisiae*. Perhaps the high level of glycosylation undertaken by *S. cerevisiae* interferes

with the *C*-mannosylation process. One of the previously attractive features of this system is the level of high-mannan glycans that *S. cerevisiae* produces. Clearly, there is an abundance of the substrate, dolichol-P-mannose that is readily available in the yeast secretory pathway. However, perhaps this abundance of glycosylation works directly against our assay design. It is possible that the high glycan levels deplete the cell of the substrate Dol-P-Man, lowering the available substrate levels such that the potential *C*-mannosyltransferases TMTC2 and TMTC3 cannot act.

For these reasons, attempts to characterize the putative CMT, TMTC2 using bacteria and yeast were abandoned and attention was shifted to studies in higher eukaryotic cells, specifically insect and mammalian cells, as will be discussed in the next chapter.

5 Eukaryotic Glycosylation Studies: *Characterization of Potential Candidates for the H. sapiens C-mannosyltransferase, Expression in Insect and Human Cell Lines*

5.1 INTRODUCTION

Protein glycosylation is a posttranslational modification that is conserved from archaea and eubacteria to eukaryotes.[90] Protein glycosylation is more complex in higher organisms, and there is considerable variation in the functions of the added glycans.[88, 87, 91] Similar to lower organisms, *N*-glycosylation assists in protein folding from yeast to humans, while *O*-glycosylation increases protein solubility and generates binding sites on cellular surfaces.[91] In humans and other higher eukaryotes, protein glycosylation and other posttranslational modifications are perhaps the most complex.[88] Complex glycosylation events underlie multiple biological processes, such as cellular growth and development, cell signaling, and protein folding and activation.[87, 350] Some glycosylation events are trivial and their losses have negligible effect on the biological viability of the organism; however, others are absolutely required for cell survival.[91,87] Such complexities in glycosylation are not yet fully understood and remain an active area of investigation.

It is interesting to note that 0.5-1% of the entire human genome is devoted to the production of proteins involved in the synthesis, degradation, and function of glycoconjugates.[351] Therefore, it is likely that glycosylation in humans is a much more involved and extensive process than is currently appreciated. Medical science has observed that loss of essential glycosylation events leads to congenital disorders of glycosylation (CDGs).[89] When proteins responsible for the glycosylation events falter or fail, there can be disastrous consequences for the developing organism. For example,

complete loss of *N*-glycosylation is lethal, while a partial loss can lead to severe multisystemic defects. [89, 352] Thus far, at least 12 *N*-glycosylation CDGs are known, and another 3 CDGs are proposed to interrupt both *N*- and *O*-glycosylation.[352] Identified defects in *O*-glycosylation include *O*-fucosylation, *O*-xylosylation, *O*-GalNAcylation and *O*-mannosylation. [352] These CDGs result in a range of defects affecting musculoskeletal development as well as the nervous system. [89, 352] Since *O*-mannosylation accounts for up to 30% of *O*-glycosylation in the human brain, severe defects in *O*-mannosylation lead to a neuronal migration disorder characterized by no psychomotor development and is usually fatal within one year.[213, 352] *O*-mannose linkages have also been identified in α-dystroglycan, which links the actin cytoskeleton to extracellular matrix in muscle and nervous tissue.[213] This finding, and the rotated abdomen mutation identified in *D. melanogaster* by Bridges and Morgan led to further research linking the loss of *O*-mannosylation to congenital muscular dystrophies. [213, 353] These are only a few examples of the severe consequences of glycosylation loss, although many CDGs remain to be identified.

Defects in glycosylation that result in lethality are not classified as CDGs. Thus there is another important glycosylation defect, the complete loss of β-*O*-GlcNAcylation, which causes loss of stem cell viability, somatic cell viability and embryonic lethality in mice.[205, 354] Further examination of *O*-GlcNAcylation has revealed multiple roles for this glycosylation event in the cell. Studies have demonstrated the dynamic interplay between protein phosphorylation and β-*O*-GlcNAcylation with ramifications in cell cycle regulation.[355] The addition and removal of *O*-GlcNAc are key to histone remodeling, transcription, proliferation, apoptosis, and proteasomal degradation, as well as the regulation of nutritive pathways in humans (insulin signaling cascade) and plants (gibberellin).[356]

In contrast to glycosylation at nitrogen and oxygen, very little is known about glycosylation at carbon. Considering the biological importance of protein glycosylation, the physiological role of *C*-mannosyl transfer and the pathological consequences of its absence are of particular interest. This novel protein posttranslational modification was identified in 1994; however, the function, purpose, and mechanism of the addition remains poorly understood.[229, 234] To obtain further information regarding this unique means of protein glycosylation, it is necessary to identify and characterize candidate proteins representing potential *C*-mannosyltransferases.

To identify potential candidates for the *C*-mannosyltransferase, we examined what is known about the protein(s) and its (their) activity. Previous research had shown that the protein is located in the endoplasmic reticulum and uses dolichol phosphate mannose as a substrate.[126] Therefore, we reasoned that our target protein should have a PMT (Protein Mannosyl Transferase) domain that binds to dolichol-phosphate-mannose (Dol-P-Man). This implied that we could identify potential candidate proteins *in silico* by comparing proteins of undetermined function in the NCBI protein database to amino acid sequences of known PMT domains. We used gi 34190043 (*H. sapiens* protein *O*-mannosyltransferase 1, or POMT-1) as the query, and subjected the relevant amino acid sequence to pblast using the NCBI "Blast" website.[357] Using all possible protein builds in the database, 22 proteins were identified as having some homology to POMT-1. Twenty one (95%) of these proteins were identified as POMT-1 splice variants, POMT-2, or fragments of either of these two proteins.

The remaining protein, known as TMTC2, that was identified during this database search did not correspond to the known POMT domain architecture. POMT proteins are known to consist of a series of distinct domains. POMTs have at the *N*-terminus of the protein a PMT followed by a series of three MIR (Mannosyltransferase, Inositol,

Ryanodine receptor) domains.[225] In contrast, TMTC2 has at the *N*-terminus a non-canonical PMT domain followed by a DUF (Domain of Unknown Function), whereas a series of tetratricopeptide (TPR) repeats is found at the *C*-terminus. Clearly, this protein exhibiting low homology to POMT-1 is a likely candidate for further investigation. Thus, we focused our efforts on TMTC2 and a close relative, TMTC3. The protein TMTC3 was not identified as a CMT candidate during the BLAST search most likely due to its non-canonical PMT domain, which exhibits lower homology to the canonical PMT domain than does the putative PMT domain of TMTC2.

Previously, in Chapter 4 we had attempted to express *tmtc2* in *E. coli* and found that the production of this protein was toxic to cells, causing cell death within 3 h of induction. We tried over 50 conditions and were unable to achieve expression of the protein; thus, we turned our attention to the *S. cerevisiae* system. *S. cerevisiae* is an excellent platform for the study of putative *C*-mannosyltransferases due to the inherent lack of an endogenous *C*-mannosylation pathway.[227] We were able to achieve expression of the genes *tmtc2* and *tmtc3* in this platform, and an *in vivo* assay was designed where *tmtc2* and *tmtc3* would be co-expressed with a gene whose product is known to be *C*-mannosylated. The gene *il-12b* was selected to be this target as the IL-12b protein is shown to be *C*-mannosylated *in vivo*.[358] Furthermore, as *C*-mannosylation occurs in the endoplasmic reticulum, the IL-12b protein, which passes through the endoplasmic reticulum, should be available for the putative *C*-mannosyltransferase. The success of this assay depends on the expressed IL-12b protein needs to be targeted to the yeast secretory pathway in amounts detectable following purification from the culture media.[126] While production of both TMTC2 and TMTC3 in *S. cerevisiae* was successful, we were not able to achieve co-expression of the *il-12b* gene. Further efforts to apply this assay to insect cells will be discussed in this chapter.

As an alternative to IL-12b, we also considered sGP from *Zaire ebolavirus* as a target protein for our *in vivo* *C*-mannosylation assays. One reason for selecting sGP is that it binds to the Fc gamma receptor III on neutrophils, but has no known activity towards any other cell type.[359] Therefore, the production of this protein is expected to be essentially benign towards insect and human cells in culture. This is in contrast to IL-12b, which has been shown to have anti-tumor activity and may present difficulties in generating sufficient quantities of protein in mammalian culture cells.[360]

sGP is a secreted glycoprotein found in the serum of individuals infected with Ebola hemorrhagic fever (EHF) and was found to be *C*-mannosylated in 2007.[230] The sGP protein is encoded by the fourth viral gene of Ebola virus, which is responsible for the production of pre-sGP, and $preGP_{1,2}$.[230] To generate sGP, transcriptional editing at the RNA level introduces an additional adenine residue that alters the last 28 amino acids of the protein *C*-terminus and introduces a stop codon to truncate the protein.[361] The edited RNA transcript is then translated to pre-sGP and directed into the secretory pathway where it undergoes multiple co- and post-translational modifications.[362] The resulting homodimeric protein is *N*-glycosylated at 5 of the 6 possible sites, carries 2 external and 2 internal disulfide bonds, and is *C*-mannosylated at W288, which is the 53rd residue from the *C*-terminus.[230, 361, 362]

This chapter describes our efforts to develop both *in vitro* and *in vivo* assays to detect *C*-mannosylation. Herein, we detail efforts to generate TMTC2 and TMTC3 in *S. frugiperda* insect cells for use in microsomal assays. We also attempt to silence the *tmtc2* and *tmtc3* genes in human cells with an endogenous *C*-mannosylation pathway for use in microsomal and *in vivo* assays as further evidence of the function of these gene products as CMTs.

5.2 EXPERIMENTAL DETAILS

5.2.1 MATERIALS

5.2.1.1 Bacterial Strains and Cell Culture lines

Escherichia coli strain DH5α was purchased from Bethesda Research Laboratories (Gaithersburg, MD). *S. frugiperda* Sf9 and Sf21 cell lines were a kind gift of Dr. Tanya Paull. *H. sapiens* HEK 293-T cell line was a kind gift of Dr. Maria Croyle.

5.2.1.2 Biochemicals

Enzymes and molecular weight standards used for the molecular cloning experiments were products of Invitrogen (Carlsbad, CA) or New England Biolabs (Beverly, MA). Agarose for DNA electrophoresis was obtained from Fisher Scientific (Pittsburgh, PA). Restriction digestion enzymes, calf intestinal alkaline phosphatase (CIP), 100X BSA, T4 ligase, and their respective buffers were products of New England Biolabs (Beverly, MA). Ni-NTA agarose and kits for DNA gel extraction, spin miniprep, and maxiprep were obtained from Qiagen (Valencia, CA). The growth media components for bacteria and yeast were acquired from Becton Dickinson (Sparks, MD). Antibiotics and chemicals such as isopropyl-β-D-thiogalactopyranoside (IPTG) and dithiothreitol (DTT) were products of Sigma-Aldrich Chemicals, Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA). All reagents for SDS-PAGE were purchased from Bio-Rad (Hercules, CA), with the exception of the prestained protein molecular weight markers, which were ordered from New England Biolabs. Oligonucleotide primers for

cloning were prepared by Invitrogen (Carlsbad, CA). Antibodies used for His_6 Western blotting procedures, monoclonal anti-polyhistidine clone His-I H1029 FLAG (M2 monoclonal), and anti-mouse IgG (alkaline phosphatase conjugate) A2429 were purchased from Sigma. Antibodies for the *myc* epitope and the protein interleukin-12 were purchased from Invitrogen. Anti-sGP (MAI-21630) antibody from Pierce was a kind gift of the Croyle Lab. For Western blot detection, nitroblue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) were both obtained from Promega. Porcine sequencing grade modified trypsin (catalog number V511A, 20 μg) was also obtained from Promega. Cell culture media D-MEM and penicillin-streptomycin, were from Sigma. Fetal Bovine Serum was from Gibco, and 200 mM L-glutamine from Hyclone. Tissue culture dishes (60 × 15 mm and 150 × 25 mm) were obtained from Falcon (Fisher) and tissue culture flasks T-75 and T-175 were from Grenier Bio-One (Frichenhausen, Germany). SF-900 II SFM is a protein-free insect culture medium from Invitrogen (Carlsbad, CA). *N*-Aza-tetraacetyl-mannose was synthesized in house by Chia-I Lin, a graduate student in the Liu lab. Synthetic peptides with and without biotin label were prepared by the Core Facilities of the Institute of Cellular and Molecular Biology at the University of Texas at Austin.

5.2.1.3 Plasmids, Vectors and DNA Manipulation

PEF1/*Myc*-His(b) from Invitrogen (Carlsbad, Ca.) was a gift of the Zhang Lab. Plasmids containing genes for TMTC2, TMTC3 and IL-12b were obtained as glycerol stocks of *E. coli* obtained from OPEN Biosystems (now a part of Thermo Fisher Scientific). Plasmid containing sGP was a kind gift of the Croyle Lab. DNA gel extraction and spin miniprep kits were obtained from Qiagen (Valencia, CA). *Pfu* DNA

polymerase was purchased from Stratagene. Oligonucleotide primers for cloning were prepared by Integrated DNA Technologies (Coralville, IA) or Invitrogen (Carlsbad, CA). KOD DNA polymerase was purchased from Novagen (Madison, WI). HuSH plasmids for silencing TMTC2 and TMTC3 were purchased from Origene (Rockville, MD).

5.2.1.4 Instrumentation

PCR reactions were carried out using an Eppendorf Mastercycler gradient thermal cycler from Brinkman-Eppendorf (Westbury, NY). For agarose gel electrophoresis, a mini-sub-cell GT from BioRad (Richmond, CA) was used, powered by a Fisher FB-300 electrophoresis power source. After electrophoresis, DNA bands were visualized using a Fisher FBTIV-08 UV transilluminator. Cell disruption was performed with a Fisher 550 Sonic Dismembrator, equipped with a standard sonicator horn. Homogenization of cells was accomplished with a glass 20 mL Potter-Elvejhem homogenizer equipped with both loose (high-clearance) and tight-fitting (low-clearance) pestles, which was purchased from Wheaton, (Millville, NJ). For SDS-PAGE gel electrophoresis, a mini PROTEAN II vertical system from BioRad (Richmond, CA) was used, equipped with an FB-300 or EC-1000-90 electrophoresis power source from Fisher, or E-C apparatus corporation, respectively. A mini-trans blot assembly apparatus used for Western blot, GelAir drying system for acrylamide gel preservation, and the necessary accessories and reagents were all products of BioRad. Nitrocellulose Hybond-C or Optitran BA-5 85 reinforced nitrocellulose membranes used for Western blot were from Amersham Bioscences or Whatman (Dussel, Germany), respectively.

To obtain pH values, a Corning 240 pH meter equipped with an Accumet electrode (Fisher) was used. Large scale centrifugation procedures were performed using

an Avanti J-25 unit or a J-E instrument from Beckman (Arlington Heights, IL). Microscale centrifugation procedures were carried out using an Eppendorf 5415C from Brinkmann Instruments, (Westbury, NY). Ultraviolet-visible spectra were recorded using Beckman DU-650 spectrophotometer, or in the case of Bradford Assays, Synergy HT plate reader by Biotek Instruments, equipped with a PC using KC4 software. Amicon YM-10 filtration products were purchased from Millipore (Billerica, MA). HPLC Separations used a Beckman 366 from Beckman Instruments (Fullerton, CA) equipped with an analytical C18 HPLC column from Varian (Lake Forest, CA). The ÄKTA FPLC Instrument was from Amersham Pharmacia Biosciences (Now GE Healthcare). Sephadex 200 gel filtration FPLC columns were from Pharmacia (Uppsala, Sweden). DNA and compounds were dried using a Savant Speedvac SC100 equipped with a Savant refrigerated condensation trap RT100 and Gel Pump GP100. DNA concentrations were measured using a NanoDrop ND-1000 UV-Vis instrument from Thermo Fisher Scientific (formerly NanoDrop Technologies LLC). Thick-walled polycarbonate centrifuge tubes (4 mL, 13 × 51 mm) (Model # 349622) were purchased from Beckman. For microsomal preparations, Ultracentrifuge Model TL-100 was used, equipped with a Ti-100.4 rotor both from Beckman. Insect cell lines were cultured in either an Isotemp Incubator Model #304, or a Percival Scientific incubator, model # 1-30BLL, equipped with an OS-15 Orbital Shaker (Fisher). Human cell lines were cultured in an Isotemp Incubator (Fisher) equipped with regulated CO_2 flow.

NMR spectra were acquired on either a Varian Unity 300 or 500 MHz spectrometer. Sequencing of subclones was performed using a capillary-based AB 3700 DNA analyzer by the Core Facilities of the Institute of Cellular and Molecular Biology at the University of Texas at Austin. Mass spectra were obtained by the Mass Spectrometry Core Facility in the Department of Chemistry and Biochemistry at the University of

Texas at Austin using a Waters QTOF Premier mass spectrometer operating in electrospray mode, or MALDI-MS using a CHCA matrix by or the College of Pharmacy at the University of Texas at Austin.

5.2.2 GENERAL PROCEDURES

The general methods and protocols for recombinant DNA manipulations were as described by Sambrook et al.[274] All protein purification operations were carried out at 4 ºC. Protein concentrations were determined according to Bradford using bovine serum albumin (BSA) as the standard.[272] The relative molecular mass and purity of enzyme samples were determined using SDS-polyacrylamide gel electrophoresis as described by Laemmli.[273]

5.2.3 EXPERIMENTAL PROCEDURES

5.2.3.1 Prokaryotic Strains and Growth Conditions

Escherichia coli DH5α used for routine cloning experiments was maintained on Luria-Bertani (LB) broth or LB agar, (1% NaCl, 1% tryptone, 0.5% yeast extract, pH 7.5, and in the case of plates, 2% agar). To liquid cultures, a volume of sterile 80% glycerol equal to 20% of the original culture volume was added and gently mixed to homogeneity. The mixture was then aliquoted into sterile Eppendorf tubes, flash frozen in liquid N_2 and stored at -80ºC until use. Glycerol stocks were then used in a 1:20 dilution to inoculate fresh media for plasmid preparations. *E. coli* DH10-Bac used for generation of bacmid stocks was maintained on Luria-Bertani (LB) broth or LB agar, supplemented with 50 µg/mL kanamycin, 10 µg/mL tetracycline, and 7 µg/mL gentamycin.

5.2.3.2 Eukaryotic Strains and Growth Conditions

S. frugiperda Sf9 cell lines were grown as adherent monolayers or as suspension cultures. Cells were maintained in Sf-900 II SFM media supplemented with 100 µg/mL penicillin-streptomycin solution, 10% fetal bovine serum (FBS) and 20 mM L-glutamate (Sf-900 II SFMs). For transfections, Grace's Insect Media was used to prepare reagents, and cells were grown in Sf-900-II SFM, supplemented with 20 mM L-glutamate and 100 µg/mL penicillin/streptomycin (Sf-900 II SFMu). For procedures in which proteins were harvested from the media, Cells were grown in Sf-900 II SFM media supplemented with 100 µg/mL penicillin-streptomycin solution, 10% Fetal Bovine Serum (FBS) (Sf-900 II SFMp). Cells were cultured at 27 ºC in a humidified environment using T-75 cm^2 tissue culture flasks. For protein expression, cells were cultured in 1 L Bellco long-necked conical culture flasks with 250 RPM shaking at 27 ºC. Sf-21 cell lines were maintained in Sf-900 II SFM media with no supplementation. Sf-21 cells were cultured in tissue culture dishes (150 × 25 mm) or T-75 cm^2 tissue culture flasks in a 27 ºC humidified environment. General procedures for insect cell culture followed the methods of Lynn et al. [363]

H. sapiens HEK 293-T were grown as adherent monolayers. Cells were cultured in Dulbecco's Modified Eagle Media (D-MEM)s containing 4500 µg/each of glucose, pyridoxine-HCl, and $NaHCO_3$, supplemented with: 50 µg/mL penicillin-streptomycin, 10% fetal bovine serum (FBS) and 20 mM L-glutamate. For plasmid transfections, D-MEMt containing 4500 µg/L each of glucose, pyridoxine-HCl and $NaHCO_3$, supplemented with: 100 µg/mL penicillin-streptomycin, and 20 mM L-glutamate was used. For procedures involving protein harvest from the media, D-MEM containing 4500

μg glucose/L, pyridoxine-HCl, $NaHCO_3$, supplemented with: 100 μg/mL penicillin-streptomycin, 10% fetal bovine serum (FBS) and no L-glutamate was used to culture cells. For tetraacetyl aza mannose feeding procedures, D-MEMm containing 1000 μg/L each of glucose, pyridoxine-HCl, and $NaHCO_3$, supplemented with: 100 μg/mL penicillin-streptomycin, 20 mM L-glutamate, 30 μM N-Aza-tetraacetyl-mannose was used. Cells were cultured in T-75 cm^2 tissue culture flasks at 37 ºC under a 5% CO_2 environment. After plasmid transfection, blasticidin (bsd) was added to the culture media at a final concentration of 50 μg/mL for selection of cells carrying the protein expression construct. After initial selection, the bsd concentration was reduced to 10 μg/mL (final concentration). For silencing plasmids, puromycin (pur) was used at a final concentration of 1 μg/mL for selection of cells carrying the stably integrated silencing construct.

5.2.3.3 PCR Reactions

Primers were prepared by dilution with TE buffer (50 mM Tris•HCl, pH 7.5, 1 mM EDTA, pH 8.0) to a concentration of 100 μM. This stock solution was further diluted to 10 or 20 μM with sterile water. pCR4-TOPO plasmid vectors containing each gene were obtained as *E. coli* DH5α glycerol stocks from OPEN Biosystems. Template DNA for *tmtc2*, *tmtc3* and *il-12b* genes was prepared by purifying the plasmid DNA from the *E. coli* DH5α strains using the Qiagen miniprep kit following the manufacturer's protocol. Polymerase chain reactions (PCR) primers and reaction programs vary for each gene and will be discussed in the appropriate sections of this chapter.

5.2.3.4 PCR Product Purification

Typically, PCR products were purified by agarose gel electrophoresis employing a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide (EtBr). The samples were diluted to 80% with 6X DNA loading dye (0.25% xylene cyanol, 0.25% bromophenol blue, 15% Ficoll 400) and electrophoresed at 100 V for 30 min. The DNA bands were visualized using an ultraviolet illumination source. DNA bands of the correct base pair length were excised from the gel with a scalpel and purified using the Qiagen gel extraction kit, following the manufacturer's protocols. However, gel slices were heated at 65 ºC, rather than 55 ºC as per instructions, until the gel was fully melted. DNA elution buffer (EB) was also heated to 65 ºC during extraction, and the elution volume was 30 µL.

5.2.3.5 Shuttle Vector Cloning into pGEM-T easy

PCR products were ligated into shuttle or expression vectors specific to each gene as described in later sections of this chapter. For ligations into the pGEM-T easy vector, the DNA was first A-tailed. Briefly, 20 µL of purified DNA from the previous step was transferred to a 0.2 µL thin-walled PCR tube. To this was added 2.5 µL of 10X Taq polymerase buffer, 2 µL of a dNTP solution and 0.5 µL of Taq polymerase. The reaction was incubated without cycling for 10-30 min at 72 ºC. The A-tailed product was purified using the Qiagen gel purification kit following the modifications for purification of PCR products, and ligated into the shuttle vector pGEM-T easy. A typical ligation mixture included 2.5-3.0 µL of purified A-tailed PCR product, 0.2-0.5 µL of the pGEM-T easy vector, 2-2.5 µL of 5X T4 ligase buffer, and 0.5 µL of T4 ligase, for a total reaction volume of 6 µL. The ligation mixture was incubated at room temperature overnight at 4 ºC. The ligation mixture (3 µL) was then mixed with competent *E. coli* DH5α (50 µL)

and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, followed by incubation for 2 min on ice. LB broth (500 µL) was added to the transformed cells and the cells were pelleted by centrifugation at 12,000 *g* for 10 s, discarding the supernatant. Cells were resuspended in 50 µL fresh LB media and plated on LB-agar supplemented with 100 µg/mL ampicillin, 100 µL of 10 mM IPTG and 100 µM of 2% X-Gal in dimethyl formamide (DMF). Positive transformants were selected after overnight incubation at 37 ºC. Plates were subsequently incubated for 8-12 h at 4 ºC to allow development of the blue color. White colonies were selected for future screening steps.

5.2.3.6 Shuttle Vector Cloning into pCR-Blunt

PCR products were ligated into shuttle or expression vectors specific to each gene as described in later sections of this chapter. In the case of pCR-Blunt, a typical ligation mixture included 3.5 µL of gel purified PCR product, 0.5 µL of pCR-Blunt vector, 2 µL of 5X T4 ligase buffer solution, 3.5 µL of sterile water, and 0.5 µL of T4 ligase, for a total reaction volume of 10 µL. The ligation mixture was incubated at room temperature for 30 to 60 min or overnight at 4 ºC. The ligation mixture (3-5 µL) was then mixed with competent *E. coli* DH5α (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, followed by incubation for 2 min on ice. LB broth (500 µL) was added to the transformed cells and incubated for 30-45 min at 37 ºC, with 200 RPM shaking. The cells were plated on LB-agar plates containing 50 µg/mL kanamycin. Positive transformants were selected after overnight incubation at 37 ºC.

5.2.3.7 Direct Cloning into Expression Vector

A Nanodrop-1000 instrument was use to quantify DNA solutions. A typical ligation mixture included 90 fmol ends of purified PCR product, 30 fmol ends of plasmid vector, 2 µL of T4 ligase buffer solution, 0.2-0.5 µL of T4 ligase, and sterile water to a volume of 10 µL. The ligation mixture was incubated at room temperature for 30 to 60 min, or overnight at 4 ºC. The ligation mixture (3-5 µL) was then mixed with competent *E. coli* DH5α (50 µL) and incubated on ice for 30 min. After incubation, the cells were transformed by heat shock at 42 ºC for 30 s, followed by incubation for 2 min on ice. LB broth (500 µL) was added to the transformed cells and incubated for 30-45 min at 37 ºC, with 200 RPM shaking. The cells were plated on LB-agar plates containing selective antibiotic. Positive transformants were selected after overnight incubation at 37 ºC.

5.2.3.8 Alkaline Lysis Procedure

Colonies appearing after overnight incubation were screened for the presence of plasmid containing PCR product using a modified alkaline lysis procedure.[277] Briefly, 2-5 mL cultures of each positive transformant were grown overnight at 37 ºC in LB or 2X YT media (1.6% tryptone, 1% yeast extract, 0.5 % NaCl, pH 7.0) in the presence of selective antibiotic. Culture volume was dependent on the copy number of the plasmid. Cultures were centrifuged at 12,000 *g* for 30 s, the supernatant was removed and the cells were resuspended in 100 µL of cold resuspension solution (50 mM glucose, 25 mM Tris•HCl, pH 8.0, 10 mM EDTA, 100 µg/mL RNase). Cells were disrupted with the addition of 200 µL fresh lysis solution (200 mM NaOH, 1% SDS), and inverted five times to mix. The lysis reaction was quenched with the addition of 150 µL of neutralization solution (3 M potassium acetate, 11.5% acetic acid). Reactions were mixed to ensure neutralization and incubated at 4 ºC for 5 min, followed by centrifugation for 5 min at 16,100 *g* at 4 ºC to

pellet cellular debris. Each supernatant containing plasmid DNA was transferred to a fresh tube and the DNA was precipitated by the addition of 2.5 volumes of 100% ethanol or one volume of isopropanol and 1/10 volume of 3 M sodium acetate pH 5.2. Each solution was mixed to ensure homogeneity and centrifuged at 16,100 *g* for 15 min at 4 ºC. The supernatant from each sample was then removed by aspiration, taking care not to disturb the DNA pellet. Each DNA pellet was washed with 70% ethanol and centrifuged again at 16,100 *g* for 5 min. The final ethanol wash was removed and each pellet was dried using a centrifugal vacuum apparatus for 2 min at room temperature to remove any remaining ethanol. Each pellet containing plasmid DNA was resuspended in 30 µL of 50 mM Tris•HCl (pH 8.5).

5.2.3.9 DNA Analytical Restriction Digestion and Fragment Analysis

Plasmid DNA obtained from the alkaline lysis preparation was then subjected to enzymatic digestion with the appropriate restriction enzymes. A typical reaction mixture contained 5 µL of DNA, 2 µL of 10X NEB reaction buffer, 0.2 µL each enzyme, and 13 µL of sterile water, for a final volume of 20 µL. In the case of shuttle vectors pGEM-T easy or pCR-Blunt, a typical reaction mixture contained 3 µL of DNA, 4 µL of 10X *Eco*RI buffer, 0.5 µL of *Eco*RI, and 13 µL of sterile water, for a final volume of 20 µL. Each reaction was incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis. Samples were electrophoresed for 30 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. For digestion products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragment. Electrophoretic results as visualized on agarose gels were used to determine the presence

of the DNA insert of the appropriate length. Colonies containing the vector with the ligated PCR product were grown at 37 ºC in LB media supplemented with selective antibiotic. After overnight incubation, cells were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the pelleted bacteria using the Qiagen miniprep kit, following the manufacturer's protocols with the following adjustments: the optional 500 µL PB wash was always performed to remove additional contaminants, and elution buffer EB was heated to 65 ºC before DNA elution.

5.2.3.10 DNA Sequence Analysis

The purified plasmid was sequenced to confirm the identity of the PCR product. For these genes, sequencing reactions using external and internal primers were submitted to ensure full sequence coverage. Specific primers used for each sequencing reaction are provided in later sections of this chapter. PHRED sequences obtained from the DNA facility were aligned against the known gene sequence. Contigs consisting of results from DNA sequencing reactions were assembled using the Vector NTI software program and compared to the original gene sequence. Both forward and reverse sequences were examined to determine that the correct gene sequence was obtained. When necessary, point mutations in the amplified sequences were compared to known Single Nucleotide Polymorphisms (SNPs) to determine if any of these is a result of mutations in the DNA template.

5.2.3.11 Secondary Cloning into Expression Vectors

When shuttle vectors were used, the PCR insert was isolated after sequencing and ligated into a plasmid expression vector. The general protocol used to generate expression constructs is described below.

STEP 1: GROWTH OF E. COLI CELLS. One colony of *E. coli* DH5α containing the gene of interest in a shuttle vector was used to inoculate 5 mL of LB media supplemented with the appropriate selective antibiotic. Cultures were grown overnight at 37 ºC with 200 RPM shaking and the plasmids were isolated using the Qiagen miniprep kit following the manufacturer's protocols with additions and alterations as previously detailed.

STEP 2: SHUTTLE VECTOR DIGESTION. Plasmid DNA obtained from this preparation was subjected to enzymatic digestion overnight with the appropriate enzymes. A typical preparatory digestion reaction contained: 30 µL of DNA, 4 µL of NEB 10X buffer, 2 µL of each enzyme and 2 µL of sterile water for a final reaction volume of 40 µL. For *tmtc*2 and *tmtc*3 pCR-Blunt constructs, a typical reaction mixture consisted of 10 µL of DNA, 5 µL of 10X NEBuffer, 35 µL of sterile water and 1 µL of each enzyme. Reactions were incubated overnight at 37 ºC. Digestion reactions were subjected to agarose gel electrophoresis for 30-40 min at 100 V using a 0.8% TAE-agarose gel with 5 µg/mL ethidium bromide. For digestion products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragment. Electrophoresis was required to separate the shuttle vector from the gene of interest.

The gel band containing the insert of appropriate size was excised with a clean, sharp scalpel and purified using the Qiagen gel purification kit, following the manufacturer's protocols with the additions and alterations as previously described.

Step 3: Expression Vector Purification and Digestion. Expression vectors were prepared by enzymatic digestion with the appropriate enzymes overnight at 37 ºC, followed by the addition of 0.1 µL CIP. The reaction was incubated for 1 h at 37 ºC. When NEBuffer 2 was not used in the digestion reaction, 0.25 volumes of 1 M Tris•HCl, pH 8 was added to the reaction. To purify digested DNA, the Qiagen gel purification kit was employed according to manufacturer's instructions for purifying PCR products.

Step 4: DNA Ligation and Transformation into E. coli. A Nanodrop instrument was used to determine the concentration of both vector and gene insert solutions. Ligations were performed at room temperature for 1-2 h or overnight at 4 ºC, with a typical reaction mixture consisting of 90 fmol ends of gene to be inserted, 30 fmol ends of vector, 2 µL of 5X ligation buffer, 0.5 µL of T4 ligase, and sterile water to a final volume of 10 µL. A total of 5 µL of the ligation reaction was incubated with 50 µL of competent *E. coli* DH5α on ice for 30 min. The transformation procedure is identical to that previously described; however, when ampicillin was the selective antibiotic used, the recovery incubation was omitted and the transformation was plated directly onto LB-agar plates with 100 µg/mL ampicillin.

Step 5: Ligation and Selection Analysis. After overnight incubation at 37 ºC, colonies appearing on the selection plates were screened for successful ligation by alkaline lysis and enzymatic digestion following the method described above with the following adjustments.[277] Digestion was carried out using the restriction enzymes specific for each construct, as indicated in the following chapter. A typical reaction mixture contained 5 µL of DNA, 2 µL of 10X NEB reaction buffer, 0.2 µL of each enzyme, and 13 µL of sterile water, for a final volume of 20 µL. Each reaction was incubated for 2 h at 37 ºC, after which time 10 µL aliquots were removed and analyzed by gel electrophoresis. Samples were electrophoresed for 30 min at 100 V using a 0.8% TAE-

agarose gel with 5 µg/mL ethidium bromide. For digestion products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragment.

Electrophoretic results as visualized on agarose gels were used to determine the presence of the DNA insert of the appropriate length. Colonies containing the vector with the successfully ligated gene of interest were grown at 37 ºC in LB media supplemented with the appropriate selective antibiotic. After overnight incubation, cells were harvested by centrifugation at 12,000 *g* for 30 s. Plasmid DNA was extracted from the pelleted bacteria using the Qiagen miniprep kit, following the manufacturer's protocols with the adjustments as described previously. The newly isolated gene-containing plasmids were used to transfect heterologous expression hosts.

5.2.3.12 SDS-PAGE

The relative molecular mass and purity of enzyme samples were determined using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli.[273] Proteins were electrophoresed using a discontinuous buffer system 25 mM Tris•HCl, 192 mM glycine and 0.1% SDS (pH 8.3).[273] Gels were cast using 75 mm plates, and in most cases 6 or 8% acrylamide was used for the separatory gel for proteins greater than 60 kDa and 12% for proteins smaller than 60 kDa. All stacking gel used was 4% acrylamide. Protein samples were diluted 2X with SDS-loading dye (62.5 mM Tris•HCl buffer pH 6.8), containing 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.0025% bromophenol blue, and heated to 100 ºC for 5 min. Samples were centrifuged briefly to collect insoluble material at the bottom of the tubes. The protein gels were electrophoresed using a setting of 30 mA per gel, allowing the gel to run until the loading

dye eluted from the bottom of the gel. Gels were removed from the glass plates, and stacking gel was removed and discarded. A corner of each gel was removed to preserve orientation of the gel samples. Gels were stained with Coomassie blue (2.5 g/L of Coomassie Brilliant Blue G-250 in acetic acid:water:methanol (1:4:5) by volume) until dark blue.[278] Gels were removed from staining solution, rinsed in ddH_2O, and submerged in destaining solution (ethanol:acetic acid:water in a ratio of 4:5:41 by volume).[278] Protein concentrations were determined according to Bradford using bovine serum albumin as the standard. [272]

5.2.3.13 Western Blot Procedure

Western blotting was used to detect the presence of IL-12b or FLAG and *c-myc* epitopes, or polyhistidine tags on heterologously expressed and purified proteins.[279] SDS-PAGE procedures were performed in duplicate for each protein sample. The gels were processed as described previously until the staining step, at which point gels were assembled into the transfer cassettes. The cassettes were transblotted onto the nitrocellulose membrane at 100 V for 1 h, using a running buffer of 25 mM Tris base, 192 mM glycine, 20% v/v methanol (pH 8.3). Cassette assembly and electrophoresis was carried out according to the directions provided by BioRad. After transblotting electrophoresis, the membrane was carefully removed from the transfer cassette and the presence of the prestained marker on the blot was visually assessed to ensure transfer was complete. The membrane was transferred to a small plastic box and immersed in 5% non-fat dry milk in Tris buffered saline and Tween® (TBST) (20 mM Tris•HCl, 150 mM NaCl, 0.5% v/v Tween® 20, pH 7.5). The membrane was blocked in this solution for 1 h, with agitation, overnight at 4 ºC or at room temperature for 1 h. The solution was then

removed and the membrane rinsed thrice in TBST to remove excess blocking solution. The membrane was then immersed in primary antibody solution (15 mL TBST, 1:30,000 dilution of monoclonal anti-polyhistidine clone His-1, or a 1:5000 dilution of αIl-12b, α-myc, or α-FLAG antibodies. All antibody solutions contained 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was decanted and the membrane washed thrice with TBST to remove any excess primary antibody solution remaining on the membrane. The membrane was subsequently incubated in secondary antibody solution (15 mL TBST, 1:30,000 dilution of anti-mouse IgG [Fc specific] alkaline phosphatase conjugate, 1:1000 dilution of 20% v/v NaN_3), and incubated with agitation overnight at 4 ºC or room temperature for 1 h. After incubation, the solution was decanted and the membrane washed thrice with TBST to remove any excess secondary antibody solution remaining on the membrane. To visualize bound antibody, the membrane was immersed in 5 mL color development solution (100 mM Tris•HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) with the addition of 66 μL NBT and 33 μL BCIP. The membrane was incubated in the dark for 2-5 min or until purple bands appeared on the blot. Development was stopped by rinsing the membrane with ddH_2O. Membranes were dried and scanned within 24 h to preserve coloration.

5.2.3.14 In-Gel Tryptic Digest of Proteins

Proteins were digested in acrylamide gels following the method of Shevchenko.[328] Briefly, major protein bands were excised from Coomassie stained gel with a fresh razor blade. Gel slices were transferred to sterile Eppendorf tubes and were destained overnight at room temperature with slight agitation in fresh destaining solution. After most of the

dye was removed from the gel bands, destaining solution was removed and replaced with 500 µL of fresh destaining solution, and destaining continued for another 2 to 3 h. The destaining solution was removed and discarded, and the gel slices were dehydrated in 300 µL of HPLC grade acetonitrile. This step was repeated as necessary until the gel slices were an opaque white color. The acetonitrile was removed and discarded. The gel slices were subjected to evaporation using vacuum centrifugation for 2-3 min to evaporate any remaining acetonitrile and then treated with fresh 10 mM DTT (100 µL) for 1 h at room temperature to reduce the proteins in the gel. The sample was centrifuged briefly and the DTT solution was removed. The gel slices were then incubated at room temperature for 1 h with fresh 50 mM iodoacetamide in water (100 µL). The sample was briefly centrifuged, and the iodoacetamide solution was removed. The gel slices were washed for 10 min with 100 mM ammonium bicarbonate (200 µL) and centrifuged briefly before removing the ammonium bicarbonate solution. Gel slices were dehydrated in HPLC grade acetonitrile (300 µL). This step was repeated as necessary until the gel slices turned white, after which the acetonitrile was left on the gel slices for 5 min. The acetonitrile was removed and discarded. This step was repeated for two washes of 5 min each, and the gel was rehydrated in 100 mM ammonium bicarbonate (200 µL) for 10 min. The sample was then centrifuged briefly and the ammonium bicarbonate solution was removed. Gel slices were then dehydrated in HPLC grade acetonitrile (300 µL). This step was repeated twice as above, for a total of two 5 min washes. The gel slices were dried by evaporation using the vacuum centrifuge for 2-3 min to remove any remaining acetonitrile. Trypsin solution (20 ng/µL) was prepared by adding 1 mL of 50 mM ammonium bicarbonate to 20 µg enzyme on ice. Trypsin solution (50- 100 µL) was added to cover the gel slices completely, and this mixture was incubated on ice for 10-15 min, until the gel pieces appeared swollen. Ammonium bicarbonate (50 mM, 20 µL) was

then added to the solution and the digestion was incubated overnight at 37 ºC. The next morning, the extract was centrifuged to collect the sample and the digestion solution was removed with a gel-loading pipette tip, ensuring exclusion of any gel pieces. Protein was further extracted from the gel slices by adding 50% acetonitrile/50% water with 5% formic acid (75 µL) and incubating at room temperature for 10 min. The sample was centrifuged and the resulting solution transferred to a clean eppendorf tube. This step was repeated for a total organic extraction volume of 150 µL.

5.2.4 CLONING OF TMTC2 AND TMTC3 FOR S. FRUGIPERDA

5.2.4.1 PCR Reactions and Product Purification

The genes *tmtc2* and *tmtc3* were amplified using pESC-His constructs as template DNA. Primers for the *tmtc2* gene were designed to have the restriction enzyme recognition sites for *EcoR*I and *Xba*I incorporated in the primer sequences (5′ start primer 5′-TAA*GAATTC*ATGATTGCAGAGTTGGTGAGC-3′ and 3′ halt primer 5′-AAA*TCT-AGA*TCAGATCTTATCGTCGTCATCC-3′), which also included the *C*-terminal FLAG tag found in the pESC-His vector. Primers for the *tmtc3* gene were designed to have the restriction enzyme recognition sites for *Sma*I and *Sph*I incorporated in the primer sequences (5′ start primer 5′-TTA*CCCGGG*ATGGCTAATATTAACCTA-AAAGAAATAACC-3′ and 3′ halt primer 5′-ATT*GCATGC*TTACTCGAGGTCTTCTT-CGG-3′), which also included the *C*-terminal *myc* epitope tag found in the pESC-His vector.

Template and DNA primers were obtained and prepared for PCR as described previously. The PCR reactions consisted of the following: 140 µL sterile water, 20 µL 10X KOD buffer, 16 µL of a 2 mM dNTP solution, 12 µL of a 25 mM $MgSO_4$ solution,

10 µL of dimethyl sulfoxide (DMSO), 1.0 µL of each primer (5′ and 3′) at 10 µM, 1 µL of template DNA, for a 200 µL total volume. The mixture was divided into two 0.5 mL thin-walled PCR tubes, and 2 µL KOD polymerase was added.

PCR was carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial polymerase activation step of 2 min at 95 ºC. The program consisted of a 20 s denaturing step at 95 ºC, an annealing step at 43.6 ºC for 10 s, an elongation step of 90 s at 70 ºC. This cycle was repeated thirty times, and the program ended with a 10 min extension at 70 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

The PCR products were purified by agarose gel electrophoresis and ligated into pCR-Blunt as described in sections 5.2.3.4 and 5.2.3.6 respectively, with the following alterations to the protocol. Ligation mixtures were incubated at 4 ºC overnight. The amount of the ligation mixture was used for the bacterial transformation was 5 µL. Colonies appearing after overnight incubation were then screened for the presence of plasmid containing PCR product using a modified alkaline lysis procedure, with the following modifications.[277] Culture volume was 3 mL and the media used was 2X YT media supplemented with 50 µg/mL kanamycin. Plasmid DNA obtained from this preparation was subjected to enzymatic digestion with *EcoR*I and analyzed as described previously. For PCR products greater than 2 kB, the duration of the electrophoretic separation was increased to allow greater resolution of digested vector and the resulting DNA fragment.

Electrophoretic results as visualized on agarose gels were used to confirm the presence of the DNA insert of the appropriate length. Colonies determined to contain the vector and successfully ligated PCR product were grown overnight at 37 ºC with 200

RPM shaking in 2X YT media supplemented with 50 µg/mL kanamycin, and after overnight incubation, plasmid DNA was isolated using the Qiagen miniprep kit as described previously. The purified plasmid was sequenced using two external primers (M13 forward and reverse) from the cloning vector and three internal primers. PHRED sequences obtained from the DNA facility were aligned against the known gene sequence.

Once the DNA sequences were confirmed to be the desired genes, *tmtc2* and *tmtc3*, they were transferred from shuttle vectors to expression vectors following the protocols in section 5.2.3.11. Briefly, the shuttle vector constructs were digested with the appropriate enzymes (*EcoR*I, *Xba*I for *tmtc2* and *Sph*I, *Sma*I for *tmtc3*). Expression vectors were amplified using DH5α cells grown in LB broth supplemented with 100 µg/mL ampicillin. The pFastBac Dual plasmids were isolated using the Qiagen miniprep kit and prepared by enzymatic digestion with the corresponding enzymes (as listed above). Plasmid vector purification, ligation and transformation procedures are identical to those previously listed, with the adjustments listed for ampicillin selection.

After overnight incubation, colonies appearing on the plates were grown in 2X YT media and screened using the modified alkaline lysis procedure in the previous section 5.2.3.8, followed by enzymatic digestion using *EcoR*I, *EcoRV* for *tmtc2* and *Sph*I, *Sma*I for *tmtc3* and analysis employing agarose gel electrophoresis. Colonies positive for vector and insert were grown in 3 mL 2X YT media supplemented with 100 µg/mL ampicillin for 6 h at 37 ºC with 200 RPM shaking. The plasmids were extracted using the Qiagen miniprep kit with modifications to the protocol as previously described. Plasmids generated via this method are TMTC2-pFastBac-Dual and TMTC3-pFastBac-Dual.

To generate TMTC2/TMTC3-pFastBac-Dual, the latter plasmid was again digested with *EcoR*I, *Xba*I and prepared as described above. The digested, CIP treated

vector was gel purified and subjected to ligation with *tmtc2*, also digested with the *EcoR*I, *Xba*I enzymes. Ligation, alkaline lysis, electrophoresis and final plasmid preparation were repeated as above to construct this plasmid. After construction, the dual expression vector containing both *tmtc2* and *tmtc3* was submitted to DNA sequencing to ensure that both genes were present in the vector.

5.2.4.2 Construction of tmtc2 and tmtc3 Bacmids

Once *tmtc2* and *tmtc3* genes were successfully ligated into the pFastBac Dual vector, the three expression plasmids were converted into bacmids.

STEP 1: PLASMID TRANSFORMATION INTO E. COLI. Chemically competent DH10Bac cells were transformed with each construct, TMTC2-pFastBac-Dual, TMTC3-pFastBac-Dual, TMTC2/TMTC3-pFastBac-Dual following the transformation procedure as described previously with the following alterations. After the heat shock step, 900 µL of 2X YT was added to the transformants and cells were allowed to recover for 4 h at 37 ºC with 200 RPM shaking. After the 4 h recovery time, each transformation was divided into aliquots and plated onto LB-agar supplemented with 50 µg/mL kanamycin, 10 µg/mL tetracycline, 7 µg/mL gentamycin, 40 µg/mL IPTG, and 80 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal). Aliquot volumes for each transformant were: 500 µL, 250 µL, 125 µL, 75 µL and 50 µL, and the plates were incubated for 2 days at 37 ºC. After this time, white colonies were clearly visible, indicating the formation of recombinant bacmids. White colonies were used to inoculate fresh LB-Agar (Kan50/Tet10/Gent7/IPTG40/X-Gal80) plates and were grown for 2 days at 37 ºC.

STEP 2: PCR SCREENING OF POSITIVE BACMID TRANSFORMANTS. White colonies were screened for the presence of recombinant insert using the PCR methodology

provided by the Invitrogen Bac-to-Bac manual. Briefly, PCR reactions contained: 32 µL of sterile water, 5 µL of 10X KOD buffer, 5 µL of a 2 mM dNTP solution, 3 µL of a 25 mM $MgSO_4$ solution, 1.5 µL of each primer (M13 forward and reverse) (5′ and 3′) at 10 µM, and 1 µL of KOD polymerase for a final volume of 50 µL in a 0.2 mL thin walled PCR tube. Colonies were picked from plates using sterile pipette tips and triturated in PCR reactions to introduce template DNA. PCR screening reactions used the program listed previously for the cloning of *tmtc2* and *tmtc3*, with the following changes. The annealing temperature was set to 55 ºC, as different primers were used for this reaction. PCR reactions were analyzed by gel electrophoresis using the protocols previously listed; however, a 0.3% TAE-agarose gel was employed due to the size of the PCR product (~5 kB). Colonies positive for insert as determined by the PCR screening method were used to inoculate 5 mL of LB-Kan50-Tet10-Gnt7 media and were allowed to grow overnight at 37 ºC with 200 RPM shaking.

5.2.4.3 Bacmid Isolation for S. frugiperda Sf9 Transformations and Gene Expression

DNA was prepared from *E. coli* DH10Bac cells using the Qiagen Maxi prep kit following the manufacturer's protocol for large constructs. Briefly, 1 mL of culture from the previous step was used to inoculate 500 mL of fresh LB media containing the selective antibiotics as listed above. After 12 h growth at 37 ºC with vigorous shaking (250 RPM), the culture was centrifuged for 15 min at 4 °C, 6000 *g* and the supernatant was removed. Cell pellets were either frozen at -80 ºC or processed immediately. The pellet was resuspended in 20 mL of P1 buffer (with 100 µg/mL RNase) and cells were lysed by the addition of 20 mL of P2 buffer. The solution was mixed by inversion 4-6 times and incubated at room temperature for 5 min. The solution was neutralized by the

addition of 20 mL of P3 buffer, inverted 5-6 times to mix the solution thoroughly, and incubated on ice for 10 min. The sample was centrifuged at 4 ºC for 30 min at 20,000 *g*, and the supernatant was filtered through hydrated filter paper. The flow through was collected in a clean 50 mL centrifuge tube, and 0.6 sample volumes of isopropanol was introduced to the filtered lysate. The sample was centrifuged for 30 min at 15,000 *g*, 4 ºC, and the supernatant was carefully decanted from samples so as not to disturb the DNA pellet. The DNA pellet was washed with 5 mL of 70% ethanol and centrifuged for 30 min at 15,000 *g*, 4 ºC. The supernatant was decanted, and the DNA pellet was dried by inverting the centrifuge tubes for 2-3 min over filter paper. DNA was resuspended in 9.5 mL of buffer EX, taking care not to shear the DNA by trituration. The samples from different centrifuge tubes were combined and subjected to exonuclease digestion with the addition of 200 µL ATP-dependent exonuclease and 300 µL (100 mM) ATP solution. The samples were heated at 37 ºC for 60 min. While the sample was incubating, 10 mL of buffer QBT was introduced to the Qiagen tip 500, which was allowed to flow through by gravity to equilibrate the tip. After sample incubation, 10 mL of buffer QS was applied to the sample and the entire volume was applied immediately to the pre-equilibrated tip 500. The sample was introduced by gravity, and the tip was washed 2X with 30 mL of buffer QC. The washes were allowed to flow through by gravity and retained for later analysis. Buffer QF was warmed to 65 ºC and 15 mL was introduced to the column, and the flow-through, containing the DNA of interest was collected in a sterile 50 mL centrifuge tube. Room temperature isopropanol, (10.5 mL or 0.7 volumes) was added to the QF elution. The tube was marked to show orientation in the centrifuge rotor, and centrifuged immediately for 30 min at 4 ºC, 15,000 *g*. The glassy DNA pellet was washed with 5 mL of 70% cold ethanol to remove excess salts left in the isopropanol pellet. The sample was centrifuged as above and the supernatant was removed. The tubes were inverted over

filter paper for 5-10 min to completely dry the pellet, which was resuspended in a small volume of TE or EB buffer at 55 ºC. Bacmid concentrations were measured using a Nanodrop instrument and stored at -20 ºC until use.

5.2.5 S. FRUGIPERDA SF9 TRANSFECTION OF BACMIDS (GENERATION OF P1 VIRAL STOCK)

Sf9 cells were removed from culture flasks and counted using a standard hemacytometer and trypan blue dye to verify cell viability.[363] To remove Sf9 cells from flasks, the flasks were first incubated for 20 min at 4 ºC. After incubation, flasks were struck sharply with the open palm to dislodge weakly adherent cells and create a cell suspension in media. To count the cells, 0.2 mL of cell suspension was removed from the culture flask and diluted with 0.3 mL phosphate buffered saline. An aliquot (0.5 mL) of trypan blue solution (0.2 % w/v) was added to the dilute cell suspension and 100 µL of this solution was placed on the hemacytometer and counted using a compound microscope. A total count of all cells generates the number of cells in the flask, and counting again only those cells that are able to exclude the blue dye gives a number of viable cells. Using the equation: [(White cells/Total cells) × 100] generates a number equal to the "% viable" cells of the total cells in the culture.

Cells were found to be 95% viable and were plated at a density of 2×10^6 cells into 60 × 15 mm culture dishes and incubated overnight at 27 ºC in a humidified environment, in Sf-900-II SFMu. The following day, transfection reagents were prepared for each bacmid. Cellfectin at room temperature was mixed well, and 8 µL was added to 100 µL of unsupplemented Grace's Media. Each bacmid (500 ng or 1-2 µL) was combined with 100 µL of unsupplemented Grace's Media. The DNA and Cellfectin mixtures were subsequently combined and allowed to incubate for 15-30 min at room

temperature. The table below shows volumes used for a sample transfection, with control plates.

5.1 Sample Transfection Reactions for TMTC2 and TMTC3 baculoviral stocks

Plate Number/Sample	Cellfectin	Media	DNA	Media
#1 Negative Control	0 µL	100 µL	0 µL	100 µL
#2 Cellfectin Control	8 µL	100 µL	0 µL	100 µL
#3 Transfection TMTC2	8 µL	100 µL	1.7 µL	100 µL
#4 DNA control TMTC2	0 µL	100 µL	1.7 µL	100 µL
#5 Transfection TMTC3	8 µL	100 µL	2.1 µL	100 µL
#6 DNA control TMTC3	0 µL	100 µL	2.1 µL	100 µL
#7 Transfection TMTC2/3	8 µL	100 µL	1.5 µL	100 µL
#8 DNA control TMTC2/3	0 µL	100 µL	1.5 µL	100 µL

The following day, media was removed and cells were washed gently with sterile phosphate buffered saline (PBS). The wash was discarded and 210 µL of transfection reagent was added to each dish, and incubated for 3-5 h at 27 ºC with gentle agitation. After incubation, 2 mL of Sf-900-II SFMu was added to each culture dish. Cultures were incubated at 27 ºC for 72 h in a humidified environment, or until signs of infection were observed in cells. Late-stage infection media was harvested from the culture and centrifuged at 500 *g* for 5 min to remove cellular debris. The supernatant was transferred to a fresh 15 mL conical tube, wrapped in foil to protect viral particles from light and stored at 4 ºC until use.

5.2.6 S. FRUGIPERDA SF9 BACULOVIRUS INFECTION (GENERATION OF P2 VIRAL STOCK)

Uninfected Sf9 cells were removed from culture flasks and counted using a standard hemacytometer and trypan blue dye to verify cell viability. Cells were found to be 95% viable and were plated at a density of 2×10^6 cells/mL into 60 × 15 mm culture dishes and incubated overnight at 27 ºC in a humidified environment and Sf-900-II SFMu, for a total of eight dishes. The following day 100 µL of each viral stock (TMTC2, TMTC3 or TMTC2/3) was added to culture dishes in duplicate, such that two dishes contained each viral stock at approximately 0.1 M.O.I. (multiplicity of infection), leaving two dishes as negative controls for comparison. The infected cultures were subsequently incubated for 48 h at 27 ºC in a humidified environment, or until signs of infection were observed in cells. Late-stage infection media was harvested from cultures and centrifuged at 500 *g* for 5 min to remove cellular debris. The supernatant was transferred to a fresh 15 mL conical tube, wrapped in foil to protect viral particles from light and stored at 4 ºC until use.

5.2.7 S. FRUGIPERDA SF9 BACULOVIRUS PLAQUE ASSAYS

Cells were counted, assessed for viability and diluted to 5×10^5 cells/mL (as calculated based on trypan blue exclusion assay). The diluted solution (2 mL) was plated in 60 × 15 mm dishes to 50% confluence in Sf-900 II SFMs, and incubated overnight at 27 ºC in a humidified environment. After incubation, plates were examined to ensure 50% confluence. Plaquing media was made by melting 4% agarose in a sterile environment, and combining this agar solution with Sf-900 media (unsupplemented). This plaquing media was incubated at 40 ºC until use. Serial dilutions of each of three P2 viral stocks was prepared by diluting 0.5 mL of P2 stock into 4.5 mL of Grace's Insect

Media (unsupplemented) to create 8 dilutions (from 10^{-1} to 10^{-8}). For each stock, six plates were selected, the media removed and immediately replaced with: no virus, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7} and 10^{-8} dilutions. These cells were incubated for 1 h at room temperature with gentle agitation, after which time the viral dilutions were removed and plates were overlayed with plaquing media. Plates were incubated at 27 ºC in a humidified environment for 7-10 days until plaques (holes in the cellular monolayer) appeared. After this time, plaques were counted and titer was calculated according to the following formula.

titer (pfu/mL) = # plaques × dilution factor × (1/ mL of inoculum/ well)

5.2.8 S. FRUGIPERDA SF9 TMTC2 GENE EXPRESSION AND PROTEIN PURIFICATION

STEP 1: GROWTH OF SF9 CELLS. Sf9 cells were removed from one 175 cm^2 flask and transferred to a sterile 1 L cylindrical Bellco culture flask and 175 mL of fresh Sf-900-II SFMs was added. TMTC2-P2 viral stock was added to an M.O.I. of 0.1, or an approximate volume of 1-3 mL. Cells were grown at 27 ºC in a humidified environment with the fast shaking on an orbital shaker for 96 h. Samples (50 mL) were taken at time points: 0, 48, 72 and 96 h post-infection.

STEP 2: CRUDE EXTRACT PREPARATION. The media was harvested at each time point by centrifugation at 2,500 *g* for 15 min at 4 ºC. Cells were washed with 22.5 mL of water and then with 17 mL of buffer B (25 mM sodium phosphate, pH 8.0, with 5 mM $MgCl_2$ and 0.2% β-mercaptoethanol). The cell pellet was frozen and stored at -80 ºC until processing. The media was then concentrated using an Amicon YM-10 stirred cell apparatus at 4 ºC until the media was less than 10 mL. The media was frozen at -80 ºC and retained for gel electrophoresis.

STEP 3: TMTC2 PROTEIN ENRICHMENT. The protein TMTC2 is expressed as a *C*-terminal FLAG-tagged protein and enriched by the method of Schutzbach.[124] All procedures were carried out at 4 ºC or on ice. The cell pellet was thawed, resuspended in 750 µL buffer B, and the cells were sonically ruptured at a power setting of 4.5 for 3 × 30 s with 1 min intervals to allow cooling. After sonication, a particulate fraction was obtained by centrifugation at 16,100 *g* for 20 min (microcentrifuge maximum speed). The supernatant was discarded and the pellet was washed twice with 2 mL of buffer B, then resuspended in 375 µL of the same buffer. The suspension was diluted with 1.3 mL of buffer C, (10 mM sodium phosphate, pH 8.0, containing 0.5 mM EDTA, 10% glycerol, 0.1% Nonidet P-40, and 1.0 mM dithiothreitol). A solution of Nonidet P-40 (200 µL of 10%) was added, and the mixture was centrifuged at 16,100 *g* for 20 min. The supernatant was removed and retained as Fraction "C". Additional enzyme was obtained from the pelleted fractions, solubilized by the addition of 1.5 mL of buffer D (0.1 M Tris acetate, pH 7.5, containing 10% (v/v) glycerol, 0.2% mercaptoethanol, and 0.1% (w/v) sodium dodecyl sulfate) to the precipitate. The precipitate was dispersed by vigorous trituration followed by the addition of 200 µL of 10% Nonidet P-40. The solution was centrifuged as above, and the supernatant was removed and retained. This procedure was repeated once more to yield a third solubilized fraction. The second and third solubilized supernatant fractions were combined and retained as Fraction "D". Fractions C and D were stored at -20 ºC.

STEP 4: SDS-PAGE AND WESTERN BLOT ANALYSIS. Aliquots (15 µL) were removed from fractions C and D isolated from each time point (0-96 h), as well as media samples (0-96 h). All samples were diluted 2X with SDS sample loading dye. Media samples were boiled for 5 min at 100 ºC, while fraction C and D samples were not boiled. SDS-Page and Western Blot were carried out according to methods detailed in section 5.4.3.12

and 5.4.3.13. Western Blot employed Anti-FLAG (primary) and anti-mouse (secondary) antibodies to detect full-length TMTC2 transcript.

5.2.9 CLONING OF IL-12B FOR S. FRUGIPERDA

5.2.9.1 PCR Reactions and Product Purification

The *il-12b* gene was amplified using pCR4-TOPO constructs from Open Biosystems as template DNA. IL-12B primers were designed to have the restriction enzyme recognition sites for *Hind*III and *Xho*I incorporated in the primer sequences (5′ start primer 5′- CCG*AAGCTT*GATGTGTCACCAGCAGTTGGTCATCTCTTGGTTTT-CCCTG-3′ and 3′ halt primer 5′- TAT*CTCGAG*ACTGCAGGGCACAGATGCCCATTC-GCTCCAAGATG-3′). The PCR reactions consisted of the following: 32.8 µL of sterile water, 5 µL of 10X KOD buffer, 5 µL of a 2 mM dNTP solution, 3 µL of a 25 mM $MgSO_4$ solution, 1.5 µL of each primer (5′ and 3′) at 10 µM, 0.2 µL of template DNA, and 1 µL of KOD polymerase, for a 50 µL total reaction volume in an 0.2 mL thin walled PCR tube.

The PCR reactions were carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial polymerase activation step of 2 min at 95 ºC. The program consisted of a 20 s denaturing step at 95 ºC, an annealing step at 64.3 ºC for 10 s, an elongation step of 20 s (20 s/kb) at 70 ºC. This was repeated thirty times, with each cycle ending in a 10 min extension at 70 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

The PCR products were purified by agarose gel electrophoresis and the Qiagen gel purification kit as described previously, with the following modifications. Elution was

reduced to 20 µL volume. PCR products were A-tailed and ligated into the shuttle vector pGEM-T easy for blue-white screening as described in section 5.2.3.5.

White colonies appearing after incubation were screened for the presence of plasmid containing PCR product using a modified alkaline lysis procedure.[277] For this screening experiment, 3 mL cultures of each positive transformant were grown overnight at 37 ºC in LB media supplemented with 100 µg/mL ampicillin, and cells were processed by the protocol detailed in section 5.2.3.8. Plasmid DNA obtained from this preparation was subjected to enzymatic digestion with *EcoR*I and agarose gel electrophoresis as described in section 5.2.3.9.

Electrophoretic results as visualized on agarose gels were used to confirm the presence of the DNA insert of the appropriate length. Colonies determined to contain the vector and successfully ligated PCR product were grown overnight at 37 ºC with 200 RPM shaking in LB media supplemented with 100 µg/mL ampicillin. After overnight incubation, plasmid DNA was extracted using the Qiagen miniprep kit and submitted for sequencing to confirm the identity of the PCR product. For the *il-12b* gene, two sequencing reactions, using two external primers (M13 forward and reverse) from the cloning vector were submitted to ensure full sequence coverage.

PHRED sequences obtained from the DNA facility were aligned against the known gene sequence. Once the DNA sequence was confirmed to be the desired gene, *il-12b*, the gene was digested from the shuttle vector with the appropriate enzymes (*Hind*III, *Xho*I) and ligated into the pMIB expression vector following the protocols described previously. Ligations were then performed at 4 ºC overnight. The transformation procedure is identical to that previously listed, with the following adjustments. The transformation was not incubated for 30-45 min after the heat shock, it was instead plated directly onto LB-agar plates containing 100 µg/mL ampicillin.

After overnight incubation, colonies appearing on the plates were grown in LB media and screened using the modified alkaline lysis procedure of the previous section 5.2.3.8, followed by enzymatic digestion using *Hind*III and *Xho*I. The digestion reactions were analyzed using agarose gel electrophoresis and colonies positive for vector and insert were grown in 10 mL LB media with 100 µg/mL ampicillin for 8-10 h at 37 ºC with 200 RPM shaking. The plasmids were extracted using the Qiagen miniprep kit with modifications to the protocol as previously described. The plasmid generated via this method is pMIB/IL-12b which was used for expression of *il-12b* in insect cells and secretion of the protein into the media.

5.2.10 S. FRUGIPERDA SF9 TRANSFECTION OF PMIB-IL-12B PLASMID

Sf9 cells were removed from culture flasks and counted using a standard hemacytometer and trypan blue exclusion assay to verify cell viability. Cells were found to be 95% viable and were plated at a density of 2×10^6 cells into 60 × 15 mm culture dishes and incubated overnight at 27 ºC in a humidified environment, in Sf-900-II SFMu. The following day, transfection reagents were prepared for each plasmid. An aliquot of each plasmid (1 µg or 1-10 µL) was combined with 1000 µL of Grace's Media, unsupplemented. Cellfectin was mixed well and 20 µL was added to the DNA-Media mix. The transfection mixtures were then combined and allowed to incubate for 15-30 min at room temperature. Table 5.2 shows sample transfection conditions with controls.

Media was removed from each dish by aspiration and cells were washed gently with sterile PBS. The wash was discarded and the transfection reagent was added dropwise to each dish. The dishes were incubated for 4 h at room temperature with gentle shaking.

5.2 Sample Transfection Reactions for Sf9 pMIB/IL-12b

Plate Number/Sample	Cellfectin	Media	DNA
#1 Negative Control	0 µL	1000 µL	0 µL
#2 Cellfectin Control	20 µL	1000 µL	0 µL
#3 Transfection IL-12b #1	20 µL	1000 µL	5 µL
#4 DNA control IL-12b #1	0 µL	1000 µL	5 µL
#5 Transfection IL-12b #2	20 µL	1000 µL	7 µL
#6 DNA control IL-12b #2	0 µL	1000 µL	7 µL

After incubation, 2 mL of Sf-900-II SFM media (unsupplemented) was added to each plate and culture dishes were incubated overnight at 27 ºC in a humidified environment. The following day, media was removed and replaced with 2 mL of Sf-900-II SFMu. After 24 h, cells were removed from each culture dish (1-6) and transferred to 25 cm^2 sterile culture flasks (1-6), cells from one new dish were transferred into one new flask, to which 5 mL of Sf-900-II SFMs was added. The cells were incubated at 27 ºC in a humidified environment for an additional 24 h, and 50 µg/mL (1 µL 100X solution per mL culture) blasticidin was added. The media was changed every 72 h, and fresh blasticidin was added until cells in control samples (no DNA, no cellfectin, no treatment) died. It was noted that some cells in the transfection flasks died, but confluence steadily increased in these flasks. After 6 days, selection was stabilized and cells from transfected cultures (3 and 5) were transferred into two new 75 cm^2 flasks and grown in 15 mL of Sf-900-II SFMs supplemented with 10 µg/mL blasticidin. Cells were grown at 27 ºC in a humidified environment until 70% confluent. At this point cells were split 1:4 and cultured in 8 175 cm^2 flasks in 25 mL of media under the same conditions as above.

5.2.11 S. FRUGIPERDA SF9 IL-12B GENE EXPRESSION AND PROTEIN PURIFICATION

STEP 1: GROWTH OF SF9 CELLS. Sf9 cells expressing *il-12b* were removed from one 175 cm^2 flask and transferred to a sterile 1 L cylindrical Bellco culture flask and 225 mL of fresh Sf-900-II SFMp supplemented with 10 µg/mL blasticidin was added. Cells were grown at 27 ºC in a humidified environment with the orbital shaker on setting 4 for 48 h.

STEP 2: CRUDE EXTRACT PREPARATION. The media was harvested and centrifuged at 2,500 *g* for 15 min at 4 ºC to remove the media. The cell pellet was retained and stored at -80 ºC until processing. The media was then concentrated using an Amicon YM-10 stirred cell apparatus at 4 ºC until the media was less than 10 mL.

STEP 3: NI-NTA CHROMATOGRAPHY. The protein IL-12b is generated as a C-terminal His_6 tagged protein and purified by affinity chromatography with Ni-NTA resin, following manufacturer's instructions with 10% glycerol in all buffers. Ni-NTA resin (2 mL) was added to the concentrated media and allowed to bind at 4 ºC for 2 h on an orbitron rotational mixer. After binding, the slurry was introduced into a 1 × 10 cm column and allowed to settle by gravity. The packed column was washed with 5 column volumes (CV) of Wash buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 10% glycerol, pH 8.0). IL-12b protein was eluted with 2 mL of Elution buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 15% glycerol, pH 8.0), which was collected in two 1 mL fractions.

STEP 4: SDS-PAGE AND WESTERN BLOT ANALYSIS. Aliquots (15 µL) of concentrated media, flow-through, wash and elution fractions were removed from samples, diluted 2X with SDS sample loading dye and boiled for 5 min at 100 ºC. SDS-Page and Western Blot were carried out according to methods detailed in section 5.2.3.12 and 5.2.3.13.

5.2.12 S. FRUGIPERDA SF9 IL-12B AND TMTC2 GENE CO-EXPRESSION AND PROTEIN PURIFICATION

STEP 1: GROWTH OF SF9 CELLS. Sf9 cells expressing *il-12b* were removed from one 175 cm^2 flask and transferred to a sterile 1 L cylindrical Bellco culture flask and 175 mL of fresh Sf-900-II SFMp supplemented with 10 µg/mL blasticidin was added. TMTC2-P2 viral stock was added to an M.O.I. of > 1, (1-3 mL of viral stock). Cells were grown at 27 ºC in a humidified environment with the orbital shaker set to 4 for 96 h. Samples (50 mL) were taken at 0, 48, 72 and 96 h after infection.

STEP 2: CRUDE EXTRACT PREPARATION. The media was harvested at each time point and centrifuged at 2,500 *g* for 15 min at 4 ºC to remove the media. Cells were washed with 22.5 mL of water and then with 17 mL of buffer B (25 mM sodium phosphate, pH 8.0, with 5 mM MgCl, and 0.2% β-mercaptoethanol). The cell pellet was frozen and stored at -80 ºC until processing. Meanwhile, the media was then concentrated using an Amicon YM-10 stirred cell apparatus at 4 ºC until the media was less than 500 µL. The media was frozen at -80 ºC and retained for gel electrophoresis.

STEP 3: TMTC2 PROTEIN ENRICHMENT. The protein TMTC2 is produced as a C-terminal FLAG epitope tagged protein and enriched by the method of Schutzbach.[124] All procedures were carried out at 4 ºC or on ice. The pellet was thawed, resuspended in 750 µL buffer B, and the cells were sonically ruptured at a power setting of 4.5 for 3 × 30 s with 1 min intervals to allow cooling. After sonication, a particulate fraction was obtained by centrifugation at 16,100 *g* for 20 min (microcentrifuge maximum speed). The pellet was washed twice with 2 mL of buffer B and resuspended in 375 µL of the same buffer. The suspension was diluted with 1.3 mL of buffer C, (10 mM sodium phosphate, pH 8.0, containing 0.5 mM EDTA, 10% glycerol, 0.1% Nonidet P-40, and 1.0 mM

dithiothreitol). Nonidet P-40 (200 µL of 10%) was added, and the mixture was centrifuged at 16,100 *g* for 20 min. The supernatant was removed and retained as Fraction "C". Additional enzyme was obtained from the pelleted fractions, solubilized by the addition of 1.5 mL of buffer D (0.1 M Tris acetate, pH 7.5, containing 10% (v/v) glycerol, 0.2% mercaptoethanol, and 0.1% (w/v) sodium dodecyl sulfate) to the precipitate. The precipitate was dispersed by vigorous trituration followed by the addition of 200 µL of 10% Nonidet P-40. The solution was centrifuged as above, and the supernatant was saved. This procedure was repeated once more to yield a third solubilized fraction. The solubilized supernatant fractions were combined and retained as Fraction "D". All fractions were stored at -20 ºC.

STEP 3: IL-12B ENRICHMENT FROM MEDIA BY NI-NTA CHROMATOGRAPHY. The protein IL-12b is generated as a C-terminal His_6 tagged protein and purified by affinity chromatography with Ni-NTA resin following the manufacturer's instructions with 10% glycerol in all buffers. Ni-NTA resin (250 µL) was added to the concentrated media and allowed to bind at 4 ºC for 2 h on an orbitron rotational mixer, then centrifuged at 1,300 *g* to pellet the resin. The supernatant was removed and the resin was washed thrice with 1.5 mL of Wash buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 10% glycerol, pH 8.0). IL-12b protein was eluted with 250 µL of Elution buffer II (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 15% glycerol, pH 8.0).

STEP 4: SDS-PAGE AND WESTERN BLOT ANALYSIS. To assess IL-12b production, 15 µL samples of concentrated media, flow-through, wash, and elution fractions were removed, diluted 2X with SDS sample loading dye and boiled for 5 min at 100 ºC. SDS-Page and Western blot were carried out according to methods detailed earlier in this chapter. Western blot employed anti-IL-12b or anti-His (primary) and anti-mouse (secondary) antibodies to detect full-length IL-12b protein.

To assess TMTC2 expression, 15 μL samples of fractions C and D from each time point (0-96 h) were removed and diluted 2X with SDS sample loading dye. SDS-Page and Western Blot were carried out according to methods detailed earlier in this chapter. Western Blot employed anti-FLAG (primary) and anti-mouse (secondary) antibodies to detect full-length TMTC2 protein.

5.2.13 S. FRUGIPERDA SF21 TMTC2, TMTC3 AND TMTC2/TMTC3 GENE EXPRESSION

STEP 1: GROWTH OF SF21 CELLS. S. frugiperda SF21 cells were grown at 27 ºC in SF-900 II SFM, with no supplements. SF21 cells, 25 mL at 1×10^6 cells per mL were transferred into 40 150 mm dishes and allowed to grow overnight at 27 ºC in a humidified environment. The following day, the cells were determined to be confluent as assessed by visual examination of cultures using light microscopy. Ten dishes were infected with each of the following recombinant baculovirus P2 stocks, TMTC2, TMTC3 or TMTC2/TMTC3 generated in step 5.2.6, at M.O.I >1 and harvested 48 to 60 h after infection. Each experimental group consisted of 10 dishes, and 10 dishes were grown but remained uninfected as negative controls.

STEP 2: CRUDE EXTRACT PREPARATION. At the time of harvesting, cells were scraped from plates using a sterile scraper and transferred into 50 mL sterile conical tubes. The cells were centrifuged at 2,000 *g* for 10 min, washed once with 0.9% NaCl and immediately frozen at -80 ºC until further manipulation. A sample of the media was frozen at -80 ºC and retained for gel electrophoresis.

STEP 3: TMTC2/TMTC3 PROTEIN ENRICHMENT. The protein TMTC2 is produced as a *C*-terminal FLAG epitope tagged protein, and the protein TMTC3 is expressed as a *C*-terminal *myc* epitope tagged protein. Screening for protein expression was carried out

by protein enrichment following the method of Schutzbach.[124] All procedures were carried out at 4 ºC or on ice. The pellet was thawed, resuspended in 750 µL buffer B, and the cells were sonically ruptured at a power setting of 4.5 for 3 × 30 s with 1 min intervals to allow cooling. A particulate fraction obtained by centrifugation at 16,100 *g* for 20 min, and was washed twice with 2 mL of buffer B and resuspended in 375 µL of the same buffer. The suspension was diluted with 1.3 mL of Buffer C. Nonidet P-40 (200 µL of a 10% solution) was added, and the mixture was centrifuged at 16,100 *g* for 20 min. The supernatant was removed and retained as Fraction "C". Additional enzyme was obtained from the pelleted fractions, solubilized by the addition of 1.5 mL of Buffer D to the precipitate. The precipitate was dispersed by vigorous trituration followed by the addition of 200 µL of 10% Nonidet P-40. The solution was centrifuged as above, and the supernatant was saved. This procedure was repeated again to yield a third solubilized fraction. The solubilized supernatant fractions were combined and retained as Fraction "D". All fractions were stored at -20 ºC.

STEP 4: SDS-PAGE AND WESTERN BLOT ANALYSIS. Aliquots (15 µL) samples of: fractions C and D, as well as the media samples were removed, diluted 2X with SDS sample loading dye and loaded onto gels without boiling. SDS-PAGE and Western Blot were carried out according to methods detailed in previous sections. Western Blot employed anti-FLAG (primary), anti-*myc* (primary) and anti-mouse (secondary) antibodies to detect full-length TMTC2 and TMTC3 proteins.

5.2.14 S. FRUGIPERDA SF21 TMTC2 AND TMTC3 PROTEIN PURIFICATION VIA MICROSOMAL ENRICHMENT

Briefly, Sf21 cells were grown in Sf-900 II medium in suspension (4-5 × 10^6 cells/mL) and 1000 mL of these cultures were co-infected with baculovirus for the production of TMTC2, TMTC3 or TMTC2/3 for 72 h.

MICROSOMAL PREPARATION METHOD I. Microsomes were prepared according to the method of Tomazin.[364] The cell pellet was suspended in microsome lysis buffer I (10 mM Tris acetate (pH 7.5), 1.5 mM magnesium acetate, 1 mM DTT and protease inhibitor cocktail (1 mM PMSF, 2 µg/mL aprotinin, 2 µg/mL leupeptin, 1 µg/mL pepstatin, 1 µg/mL chymostatin)) for 10 min on ice, and the cells were disrupted using a Dounce homogenizer. Microsomal lysis buffer II (40 mM Tris acetate (pH 7.5), 250 mM sucrose, 25 mM potassium acetate, 3.5 mM magnesium acetate, 0.5 mM calcium acetate) was added and nuclei were removed by centrifugation at 800 *g* for 10 min. Mitochondria were removed by centrifugation at 10,000 *g* for 10 min, then the crude microsomes were pelleted by centrifugation at 100,000 *g* for 60 min. The membrane pellets were resuspended in microsome buffer III (50 mM Tris acetate (pH 7.5), 250 mM sucrose, 25 mM potassium acetate, 5 mM magnesium acetate, 0.5 mM calcium acetate, 1 mM DTT and protease inhibitor cocktail) and further purified by centrifugation through a 1.1 M sucrose cushion at 150,000 *g* for 5 h. The supernatant was discarded and membrane pellets were resuspended in microsome buffer III at 4-5 mg of membrane protein/mL, flash frozen in liquid N_2 and stored at -80 ºC. The protein content of these microsomal preparations was determined by the Bradford method with BSA as the standard. [272]

MICROSOMAL PREPARATION METHOD II. Microsomes were also prepared by the method of Jennewein et al.[365] The infected cells were harvested by centrifugation (700 *g* for 10 min at room temperature), and washed with 0.9% NaCl. The washed cell pellet

was then resuspended in 5 mL of 50 mM Hepes buffer (pH 7.5) containing 0.5 mM EDTA, 0.1 mM DTT, 10% (v/v) glycerol, and the cells were lysed by sonication, using the microtip probe. Cell debris was removed by centrifugation (10,000 *g* for 10 min at 4°C), and the resulting supernatant was then centrifuged at 105,000 *g* for 1 h at 4°C to provide the microsomal enzyme preparation, which was either assayed immediately or frozen for later use in liquid N_2 and stored at -80°C. The protein content of these microsomal preparations was determined by the Bradford method with BSA as standard.[272]

MICROSOMAL PREPARATION METHOD III. Microsomes were also prepared by the method of Chaney et al.[366] The infected cells were harvested by centrifugation (700 *g* for 10 min at room temperature), and washed with 0.9% NaCl. Crude membranes were prepared from washed cells resuspended in 20 mM MES buffer (pH 6.0) containing 1 mM $CaCl_2$ and a mixture of protease inhibitors: soybean trypsin inhibitor (200 µg/mL), benzamidine (5 mM), and paramercurisulfonylfluoride (PMSF; 4 mM). After swelling 15 min on ice, the cells were Dounce homogenized (75-150 strokes), centrifuged at 530 *g* for 10 min at 4 °C, and the supernatant was spun at 100,000 *g* for 1 h at 4 °C. The membrane pellet was resuspended in 100 µL 1% Triton X-100 containing 20 mM MES (pH 6.0) with 4 mM PMSF to provide the microsomal enzyme preparation, which was either assayed immediately or frozen for later use in liquid N_2 and stored at -80°C. The protein content of these microsomal preparations was determined by the Bradford method with BSA as the standard.[272].

SDS-PAGE AND WESTERN BLOT ANALYSIS. Samples (15 µl) of microsome fractions were aliquotted, diluted 2X with SDS sample loading dye and loaded onto gels without boiling. SDS-Page and Western blot were carried out according to methods detailed earlier in this chapter. Western blot employed anti-FLAG (primary) antibodies to detect

full-length TMTC2 protein and anti-*myc* (primary) antibodies to detect full-length TMTC3 protein. Cells expressing both genes were subjected to two Western blots (anti-FLAG (primary) TMTC2 and anti-*myc* (primary) TMTC3) in order to ensure expression of both genes. Anti-mouse was the secondary antibody used in all cases.

5.2.15 sGP Gene Amplification

5.2.15.1 Cloning for sGP Expression Constructs

Primers were designed to incorporate the chosen restriction enzymes into the sequence, as indicated by the italicized underlined text within primer sequences. For expression in *E. coli*, using the pET28b vector, *Nde*I and *Hind*III were selected to yield an N-terminal-His_6 protein tag on the protein (5′ primer 5′-ACGCACG*CATATG*GGCGT-TACAGGAATATTGCAGTTACCTCG-3′, 3′ primer 5′-ACG*AAGCTT*AGATCGGACA-CTGCAGCT-TCCCTTCCTTG-3′). Primers were also designed for sGP expression in mammalian cells as well. This was done by removing the stop codon from the gene sequence, using *BamH*I and *Xba*I as the restriction enzymes (5′ primer 5′-CA*GGAT-CC*ATGGGCGTTACAGGAATATTGCAGTTACCTCG-3′, 3′ primer 5′-CA*TCTAGA*T-AGATGCGACACTGCAGCTTCCCTTCCTTG-3′). This construct would yield a C-terminal His_6 tagged protein with the *myc* epitope when ligated into the pEF6/*myc*-His vector. The PCR reactions consisted of the following: 64 µL of sterile water, 10 µL of 10X KOD buffer, 10 µL of a 2 mM dNTP solution, 6 µL of a 25 mM $MgSO_4$ solution, 3 µL of each primer (5′ and 3′) at 10 µM, 2 µL of template DNA, and 2 µL of KOD polymerase, for a 100 µL total reaction volume in an 0.5 mL thin walled PCR tube.

The PCR reaction was then carried out using an Eppendorf thermocycler with the lid temperature set at 105 ºC and an initial denaturing step of 2 min at 95 ºC. The

program consisted of a 20 s denaturing step at 95 ºC, an annealing step at 61 ºC for 10 s, and an elongation step for 20 s (20 s/kb) at 70 ºC. This was repeated 30 times, with each cycle ending in a 10 min extension at 70 ºC. The temperature was reduced to 4 ºC until the product was removed from the thermocycler. PCR products were stored at 4 ºC until purified.

PCR products were purified using 0.8% TAE-agarose gel electrophoresis followed by use of the Qiagen Gel extraction kit as described previously, and ligated into *StuI* digested pCR-Blunt. The reactions were incubated at 25 ºC (room temperature) for 30 to 60 min, and 5 µL of the ligation mixture was used for the transformation. Colonies appearing on plates after overnight incubation were used to inoculate 2-3 mL cultures in LB media supplemented with 50 µg/mL kanamycin. Plasmids were extracted from cells by alkaline lysis, digested with *EcoR*I, and analyzed as described previously.[277] Samples that appeared to contain both vector and insert were cultured again, and plasmids were isolated using the Qiagen miniprep kit. The isolated plasmids were quantified, and samples of 500 ng with $A_{260/280} > 1.8$ and $A_{260/230} > 1.8$ were submitted for DNA sequencing using the M13 forward and reverse primers supplied by the University of Texas ICMB Core facility for forward and reverse sequencing, respectively.

Plasmids containing PCR products with correct sequences were then digested with the appropriate enzymes (*Nde*I, *Hind*III for pET28b, and *BamH*I, *Xba*I for pEF6/*myc*-His) and the digested inserts were gel purified as previously described. sGP genes were ligated into the pET28b and pEF6/*myc*-His vectors. Transformations containing sGP-pEF6/*myc*-His plasmid were plated directly on LB-agar plates supplemented with 100 µg/mL ampicillin, while sGP-pET28b transformants were allowed to recover for 30-45 min before selection on LB-agar plates supplemented with 50 µg/mL kanamycin. Colonies growing on these plates were then used to inoculate 5 mL

cultures grown in LB media with the appropriate selective antibiotic overnight at 37 ºC, and plasmids were isolated using the alkaline lysis procedure above. After isolation, 15 μL of the plasmid solution was digested with *Nde*I and *Hind*III in the case of pET constructs, or *BamH*I and *Xba*I in the case of pEF constructs. All digestion reactions were incubated and electrophoresed following procedures detailed previously. Colonies containing plasmids observed to have the insert and plasmid of the correct size were then grown overnight in 5 mL cultures grown in LB-Kan 50 broth (pET) or LB-Amp100 broth (pEF) at 37 ºC with 200 RPM shaking, and the sGP expression plasmids were isolated using the Qiagen miniprep kit following the manufacturer's protocols.

5.2.16 sGP Protein Expression in E. coli

Step 1: Growth of E. coli Cells. sGP-pET28b plasmid was transformed into chemically competent BL-21(DE3) Rosetta II cells using heat shock following the protocols previously described. Colonies growing on transformation plates were then used to inoculate two 25 mL cultures grown in LB-Kan 50-Cam 35 media overnight at 30 or 37 ºC with 200 RPM shaking. After overnight incubation, 6 L of fresh LB-Kan50-Cam35 were inoculated using 6 mL overnight culture per L broth. Cultures were grown at 30 ºC with 250 RPM shaking until the culture OD_{600} reached 0.6, at which point, gene expression was induced with 1 mL of 1 M IPTG, for a final concentration of 1 mM. The temperature was reduced to 25 ºC and the cultures were incubated for an additional 16-20 h. At the end of the incubation, the cultures were harvested by centrifugation at 4,000 *g* for 20 min at 4 ºC, which yielded cell pellets of approximately 45 g.

Step 2: Crude Extract Preparation. Cells were resuspended in a 5X cell weight volume of sGP buffer (50 mM Tris buffer, pH 7 at 25 ºC, 20% v/v glycerol).

Lysozyme (1 mg/mL) was added to the mixture and the cells were incubated on ice for 1 h, stirring every 15 min to disperse clumps. The cells were disrupted by sonication using pulses of 20 s (18 × 20s) followed by cooling periods of 40 s, for an average total pulse time of 6 min. Lysed cells were then centrifuged for 20 min at 4 ºC and 10,000 *g*.

STEP 3: AMMONIUM SULFATE PRECIPITATION. Supernatant was decanted and subjected to ammonium sulfate precipitation by the addition of 64 g/L of the salt (40%), while the cell lysate stirred on ice for 1 h. The precipitate, which contained sGP, was collected by centrifugation for 20 min at 4 ºC, 10,000 *g*. The pellet was then resuspended in 40 mL of sGP buffer and dialyzed against 3 L of the same buffer at 4 ºC, exchanging the buffer every hour. After dialysis, the lysate was centrifuged to remove any insoluble protein for 20 min at 4 ºC, 10,000 *g*. The supernatant was retained and resulting pellets resuspended in 40 mL of sGP buffer.

STEP 3: DEAE COLUMN PURIFICATION. Centrifugation was repeated and the supernatants from both centrifugation steps was pooled and loaded onto a 2.5 × 25 cm DEAE column equilibrated with 500 mL of sGP buffer using a peristaltic pump with the flow rate set at 1.5 mL/min. Fractions (22.5 mL) were collected, beginning as the sample was introduced to the column. After the sample was fully loaded, the column was washed with 50-100 mL of sGP buffer and a then linear gradient of sGP buffer (1 L) and sGP buffer + 1 M NaCl (1 L). The column was allowed to run overnight (for 12 h). All fractions were screened for the presence of protein using Bradford's reagent, following the manufacturer's protocols for protein microassay. Briefly, Bradford concentrate (BioRad) was diluted 5X with ddH_2O. Aliquots of this solution (200 µL) were dispensed into each well of a 96 well plate. Samples (10 µL) of each fraction were added to individual wells of the plate and the solution was allowed to develop at room temperature

for 1 h, although fractions containing protein were clearly visible after only 2-3 min of development.

STEP 4: SDS-PAGE AND WESTERN BLOT. Fractions testing positive for the presence of protein (including the flow-through fractions) were screened by SDS-PAGE and Western Blot for the presence of sGP. A sample (50 μL) of each positive fraction was diluted with an equal volume of 2X SDS buffer with 5 mM βME and boiled for 5 min at 100 ºC. One gel was subjected to Coomassie staining to visualize protein bands, and the other was subjected to Western blotting to verify the presence of our protein of interest, following the protocols listed previously. Mouse anti-sGP was the primary antibody used for Western Blot, and the secondary antibody was anti-mouse. In general, our protein of interest was found in the first 10-15 flow-through fractions, and could be clearly visualized on both the Coomassie stain and Western Blot.

5.2.17 sGP PROTEIN EXPRESSION IN HEK-293T (MAMMALIAN TISSUE CULTURE) CELLS

5.2.17.1 Preparation of HEK-293T (Mammalian Tissue Culture) Cells

sGP-pEF6/*myc*-His plasmid was transformed into HEK-293T cells following standard mammalian protocols. Briefly, low passage HEK-293T cells were grown using D-MEMs at 37 ºC, under a 5% CO_2 atmosphere until they reached approximately 70% confluence in a T-75 sterile culture flask. The media was removed and cells washed with 5 mL phosphate buffered saline (PBS). Trypsin (0.25%)-EDTA (1 mM) (2 mL), was introduced to the flask and incubated at 37 ºC under 5% CO_2 for 5 min. After incubation, the mixture was triturated to mix cells and the flasks were rinsed with this mixture to remove any cells yet adhering to the flask walls. Aliquots (200 μL) of this mixture were

transferred into each of 10 × 60 mm culture dishes. Fresh D-MEMs, (1.8 mL) media was added to the dishes and cells were incubated overnight at 37 ºC, under 5% CO_2 to allow the cells to recover and adhere to the plate.

5.2.17.2 sGP Plasmid DNA HEK-293T Transfection and Selection

Plasmids were prepared using the Qiagen miniprep kit, and purity was examined using a Nanodrop apparatus. Only high concentration (> 200 ng/µL) plasmid preparations with $A_{260/280}$ > 1.8 and $A_{260/230}$ > 1.8 were used for transfections. Serum free D-MEMt was used to dilute 8 µg of plasmid DNA to a final volume of 500 µL and mixed gently. Lipofectamine (2 mg/mL, 20 µL) was diluted with 480 µL of serum free D-MEMt and incubated at room temperature for 5 min. The transfection mixtures for each of 10 plates were prepared as shown in the following table. DNA concentrations used were 8 µg per transfection, and DNA volumes added to the transfection to achieve a concentration of 8 µg were calculated using the following formula:

8000 ng/ plasmid DNA concentration ng/µL = required volume of DNA (µL)

Four different plasmids were used for these transfections, and a DNA control was prepared for each plasmid transfection. Sample transfections are shown in table 5.3. After incubation, the DNA and lipofectamine solutions were mixed together and incubated at room temperature for 20 min. While the transfection mixture was incubating, media was removed from culture dishes and cells were gently washed with PBS buffer. The transfection mixture was added to each dish and dishes were rocked gently to ensure even coverage.

5.3 Sample Transfection Reactions for HEK-293T/pEF*myc*HIS-sGP

Plate	Lipofectamine	Media	DNA	Media
#1 Negative Control	0 µL	500 µL	0 µL	500 µL
#2 Lipofectamine Control	20 µL	480 µL	0 µL	500 µL
#3 Transfection # 1	20 µL	480 µL	20 µL	480 µL
#4 DNA control # 1	0 µL	500 µL	20 µL	480 µL
#5 Transfection # 2	20 µL	480 µL	20 µL	480 µL
#6 DNA control # 2	0 µL	500 µL	20 µL	480 µL
#7 Transfection # 3	20 µL	480 µL	23 µL	477 µL
#8 DNA control # 3	0 µL	500 µL	23 µL	477 µL
#9 Transfection # 4	20 µL	480 µL	28 µL	472 µL
#10 DNA control # 4	0 µL	500 µL	28 µL	472 µL

The dishes were incubated at 37 ºC under 5% CO_2 overnight. The next day, cells were washed with PBS, trypsinized using 0.5 mL trypsin-EDTA as above and each sample was transferred into a new 25 cm^2 culture flask, for a total of 10 flasks. Fresh D-MEMs, 4.5 mL, was added to each flask and the cells were incubated for 24 h at 37 ºC under 5% CO_2.

The following day, media was removed and 5 mL of fresh media supplemented with 10 µg/mL blasticidin (bsd) was added to each 25 cm^2 flask, and the cells were incubated at 37 ºC under 5% CO_2. After 2-3 days, cells in the controls (DNA only, LF only, or no treatment) were found to have died. Transfection reactions also experienced low levels of cellular death, as noted by a few floating cells in the media. However, this

cell death was less than 10% of the cells in the transfection mixture as estimated by remaining confluence in the culture flasks. For the most part, transfected cells were adherent and healthy in the presence of 10 µg/mL bsd.

5.2.17.3 sGP Expression in HEK-293T

After 72 h, media was removed from the transfected cells and the cells were washed with PBS. Trypsin-EDTA (0.5 mL) was introduced to each culture flask and the flasks were incubated for 5 min at 37 ºC under 5% CO_2. After incubation, cell suspensions were triturated to disperse clumps and to remove any remaining adherent cells from the flasks. Cell suspensions were transferred into 4 75 cm^2 flasks and 14.5 mL of D-MEMs was added. After overnight incubation at 37 ºC under 5% CO_2, media was replaced with 15 mL of fresh D-MEM supplemented with 10 µg/mL bsd. Cells were grown at 37 ºC under 5% CO_2 in D-MEM supplemented with 10 µg/mL bsd until they reached 70% confluence. At this point, cells were again washed with PBS and trypsinized for transfer to 8 185 cm^2 flasks. Cells were allowed to grow in 25 mL of D-MEMs supplemented with 10 µg/mL bsd at 37 ºC under 5% CO_2. The media was changed every 72-96 h and retained at 4 ºC until 500 mL was collected.

5.2.17.4 N-Aza-tetraacetyl Mannose Labeling of sGP in HEK-293T

5.2.17.5 TRIAL I

Cells expressing *sGP* were grown under standard conditions, 37 ºC, 5% CO_2, in D-MEMs . To prepare cells for ManNaz feeding, media was removed and the cells were washed with PBS. Trypsin-EDTA (1 mL) was introduced to each culture flask and the flasks were incubated for 5 min at 37 ºC under 5% CO_2. After incubation, cell

suspensions were triturated to disperse clumps and remove any remaining adherent cells from the flasks. Cell suspensions were split 1:2 into 8 75 cm^2 flasks and 14.5 mL of D-MEMs was added. After overnight incubation at 37 ºC under 5% CO_2, media was replaced with 15 mL of fresh D-MEMs supplemented with 10 µg/mL bsd. Cells were grown at 37 ºC with 5% CO_2 in D-MEMs supplemented with 10 µg/mL bsd until they reached 70% confluence. At this point, cells were again washed with PBS and 15 mL of D-MEMm was added to culture flasks. Cells were allowed to grow in the ManNaz (30 µM) labeling media supplemented with 10 µg/mL bsd at 37 ºC with 5% CO_2. The media was changed every 48-72 h and retained at 4 ºC.

5.2.17.6 TRIAL II

Cells expressing *sGP* were grown under standard conditions, 37 ºC, 5% CO_2, in D-MEMs. To prepare cells for ManNaz feeding, media was removed and the cells were washed with PBS. Trypsin-EDTA (1 mL) was introduced to each culture flask and the flasks were incubated for 5 min at 37 ºC under 5% CO_2. After incubation, cell suspensions were triturated to disperse clumps and remove any remaining adherent cells from the flasks. Cell suspensions were split 1:2 into 8 75 cm^2 flasks and 14.5 mL of D-MEMs was added. After overnight incubation at 37 ºC with 5% CO_2, media was replaced with 15 mL of fresh D-MEMs supplemented with 10 µg/mL bsd. Cells were grown at 37 ºC under 5% CO_2 in D-MEMs supplemented with 10 µg/mL bsd until they reached 70% confluence. At this point, cells were again washed with PBS and 7.5 mL D-MEMm ManNaz feeding media was added to culture flasks. The remaining volume used in culture flasks was standard D-MEMs. Cells were allowed to grow in D-MEMs/D-MEMm media (15 µM ManNaz) supplemented with 10 µg/mL bsd at 37 ºC under 5% CO_2. The media was changed every 48-72 h and retained at 4 ºC.

5.2.18 sGP Protein Purification from HEK-293T Media

5.2.18.1 Method I

Step 1: Concanavalin A Lectin Chromatography. The following purification method is adapted from Barrientos et al. and Falzarno et al.[230,362] The yellow media samples collected from 2-3 72-96 h harvests were pooled and concentrated using an Amicon stirred cell unit at 4 ºC under nitrogen to an average volume of 2-5 mL. The concentrate was introduced to 1 mL of Concanavalin A beads and the slurry was incubated overnight at 4 ºC on a rotational mixer. The slurry was loaded onto a 1 × 10 cm column and allowed to flow through by gravity. The packed beads were washed with 10 column volumes of Z-Con Wash buffer (20 mM Tris•HCl, 500 mM
NaCl, pH 7.4) and the proteins were eluted with 2 CV of Z-Con Elution buffer (20 mM Tris•HCl, 500 mM, NaCl, pH 7.4 0.75 mM methyl-α-D-glucopyranoside). All fractions were screened for the presence of protein using Bradford's reagent, following the manufacturer's protocols for protein microassay.

Step 2: SDS-PAGE and Western Blot. Fractions testing positive for the presence of protein (including the flow-through fractions) were screened by SDS-PAGE and Western Blot for the presence of sGP. Samples (20 µL) of each positive fraction were diluted with an equal volume of 2X SDS buffer with 5 mM βME and boiled for 5 min at 100 ºC. One gel was subjected to Coomassie staining to visualize protein bands, and the other was subjected to Western blotting to verify the presence of our protein of interest, following the protocols listed previously. Mouse anti-sGP was the primary antibody used for Western Blot, and the secondary antibody was anti-mouse. In general,

our protein of interest was found in the wash fractions, and could be clearly visualized on both the Coomassie stain and Western Blot.

5.2.18.2 METHOD II

STEP 2: DEAE COLUMN CHROMATOGRAPHY. The yellow media samples collected from three 72-96 h harvests were pooled and concentrated using an Amicon stirred cell unit at 4 ºC under nitrogen to an average volume of 5-10 mL. The concentrate was loaded onto a 2.5 × 25 cm DEAE column equilibrated with 500 mL of 100 mM MES (pH 5.5 at 25 ºC) buffer using a peristaltic pump with the flow rate set at 1.5 mL/min. Fractions (22.5 mL) were collected, beginning as the sample was introduced to the column. After the sample was fully loaded and loaded and washed with 250 mL of 100 mM MES (pH 5.5 at 25 ºC) buffer, the column was washed with 100 mM MES (pH 5.5 at 25 ºC) buffer + 1 M NaCl for 4 h. All fractions were screened for the presence of protein using Bradford's reagent, following the manufacturer's protocols for protein microassay.

STEP 2: SDS-PAGE AND WESTERN BLOT. Fractions testing positive for the presence of protein (including the flow-through fractions) were screened by SDS-PAGE and Western Blot for the presence of sGP. Samples (50 µL) of each positive fraction were diluted with an equal volume of 2X SDS buffer with 5 mM βME and boiled for 5 min at 100 ºC. One gel was subjected to Coomassie staining to visualize protein bands, and the other was subjected to Western blotting to verify the presence of our protein of interest, following the protocols listed previously. Mouse anti-sGP was the primary antibody used for Western Blot, and the secondary antibody was anti-mouse. In general, our protein of interest was found in the first 2-4 flow-through fractions, and could be clearly visualized on both the Coomassie stain and Western Blot.

STEP 3: FPLC SEPHADEX 200 COLUMN CHROMATOGRAPHY. The fractions identified as containing sGP from the previous step were further purified by Fast-Protein Liquid Chromatography (FPLC). All FPLC separations were performed using a Sephadex 200 column with 500 µL of sample for each injection. The sample was eluted using an isocratic wash with 100 mM MES buffer, pH 5.5 as solvent A over two column volumes, or 46 mL. The flow rate was 0.5 mL/min and the detector was set at 280 nm. The sGP protein eluted in a range from approximately 12 to 16 mL. Aliquots of each fraction of interest were taken for analysis, while the fractions themselves were immediately flash frozen in liquid nitrogen and stored at -80 ºC.

STEP 4: SDS-PAGE AND WESTERN BLOT. FPLC fractions containing protein, as indicated by A_{280}, were screened by SDS-PAGE for the presence of sGP. Samples (50 µL) of each positive fraction were diluted with an equal volume of 2X SDS buffer with 5 mM βME and boiled for 5 min at 100 ºC. One gel was subjected to Coomassie staining to visualize protein bands, and the other was subjected to Western blotting to verify the presence of our protein of interest, following the protocols listed previously. Mouse anti-sGP was the primary antibody used for Western blot, and the secondary antibody was anti-mouse. In general, our protein of interest eluted in the first or second protein peak ten flow-through fractions under these gradient conditions, and could be clearly visualized on both the Coomassie stain and Western blot.

5.2.18.3 sGP Protein In-Gel and Solution Digestion

Bands observed to contain our protein by Western blot were excised from the corresponding Coomassie stained gel and prepared for MS analysis following the protocols described previously with the following additions and adjustments. Gel slices

were incubated in 100 µL of fresh 10 mM DTT for 2 h at 65 ºC. Samples were also incubated in 100 µL of fresh 50 mM iodoacetamide solution for 2 h at 65 ºC.

A solution of sGP protein (approximately 500 pmol) was dialyzed against 50 mM Tris•HCl buffer (pH 8.0) containing 8 M urea. dithiothreitol (DTT) was added to a final concentration of 10 mM, and the sample was incubated for 4 h at room temperature. The sample was then treated with 25 mM iodoacetamide for 30 min at room temperature. The reaction was quenched with excess DTT, and the sample was immediately dialyzed against 25 mM Tris•HCl buffer (pH 8.0). This generates reduced carboxymethylated [RCM]sGP.

To prepare tryptic peptides for MS analysis, 500 pmol of [RCM]sGP in 25 mM Tris•HCl (pH 8.0) was treated at 37 °C with trypsin (enzyme to substrate ratio of 1:100 wt/wt). The digestion was allowed to proceed overnight and then stopped by freezing the sample at -20 ºC. Both samples were submitted for MS analysis using MALDI and ESI ionization methods.

5.2.18.4 HPLC Purification of sGP Trypsin Digested Peptides

Solution and in-gel digests of sGP protein were both subjected to HPLC for additional purification of the target peptide VNPEIDTTIGE***W****AFW*ETK(K). After overnight incubation, the samples were filtered through a YM-10 centrifugal filter at 14,000 *g* for 20 min at room temperature. The peptide was purified from each sample using HPLC, equipped with a reverse phase C-18 column. Eluent A was water, and eluent B was 90% acetonitrile, with 0.1 % TFA. The gradient was linear from 100% eluent A with increasing eluent B at 1%/min, and samples were monitored at 280 nm (tryptophan). Peaks with absorbance at 280 nm were collected and lyophilized to

concentrate fractions. The concentrated fractions were submitted for MALDI-MS analysis.

5.2.19 MICROSOMAL IN VITRO C-MANNOSYLATION ASSAYS

The peptides N-AC-KPPQFAWAQWFE-NH_2 and N-AC-K^{biotin}PPQFAWAQWF-NH_2 designed based on the RNAse2 sequence, with Thr-6 replaced by Ala to prevent possible *O*-glycosylation were selected as the mannosyl acceptor substrates.[126] The peptide was incubated in the presence of microsomes from rat liver, yeast and insect cells, followed by HPLC purification using a reverse phase C-18 column (acetonitrile/TFA gradient). To ease purification, biotin labeled peptides were used. Briefly, 400 µg peptide (0.9 mM), 15 µL of microsomal protein (125 µg/µL), 0.2% v/v Triton X-100 and protease inhibitors (2 µg/mL benzamidine, 5 µg/mL pepstatin A, 5 µg/mL leupeptin, 2 mM EDTA) in HKA buffer (20 mM Hepes pH 7.2, 110 mM potassium acetate) in a final volume of 240 µL. Assays were incubated overnight at 37 ºC.

PURIFICATION OF NON-LABELED PEPTIDE. After overnight incubation, the samples were filtered through a YM-10 centrifugal filter at 14,000 *g* for 20 min at room temperature. The peptide was purified from each sample using HPLC, equipped with a reverse phase C-18 column. Eluent A was water, and eluent B was 90% acetonitrile, with 0.1 % TFA. The gradient was linear, increasing eluent B at 1%/min, and samples were monitored at 280 (tryptophan abs). Fractions displaying absorption at 280 nm were collected and concentrated by lyophilization. The concentrated fractions were submitted for MALDI-MS analysis.

PURIFICATION OF BIOTIN LABELED PEPTIDE. After overnight incubation, 50 µL of streptavidin bead suspension was added to each sample. Slurry was mixed with gentle rocking for 1 h at room temperature and centrifuged at 6,000 *g* to pellet the beads. Supernatant was removed by aspiration and retained. The beads were washed in HKA buffer twice. Briefly, beads were resuspended in 200 µL of HKA buffer and mixed well, then centrifuged to pellet the beads. The HKA wash was removed and this step was repeated for a total of three washes. The peptide was removed from beads by gently agitating the beads at room temperature for 5-10 min in a solution of 90% acetonitrile and 0.5% TFA (pH ~2). The beads were again pelleted by centrifugation and the supernatant containing peptide was transferred to a fresh eppendorf tube and concentrated to a volume of 25 µL. These samples were submitted for MALDI-MS analysis.

5.2.20 SILENCING PLASMID HEK-293T TRANSFECTION AND SELECTION

Plasmids were diluted with TE buffer to a concentration of 1 µg/mL. Serum free D-MEMt was used to dilute 1 µg of plasmid DNA to a final volume of 500 µL and mixed gently. Lipofectamine (2 mg/mL, 20 µL) was diluted with 480 µL of serum free D-MEMt and incubated at room temperature for 5 min. The transfection mixtures for each of the ten plates were prepared as shown in the following table. Four different plasmids were used for these transfections, and a DNA control was prepared for each plasmid transfection. The sample transfections are shown in table 5.4.

After incubation, the DNA and lipofectamine solutions were mixed together and incubated at room temperature for 20 min. While the transfection mixture was incubating, media was removed from culture dishes and cells were gently washed with PBS buffer.

5.4 Sample Transfection Reactions for HEK-293T/Hush Vectors

Plate	Lipofectamine	Media	DNA	Media
#1 Transfection tmtc2 #1	20 µL	480 µL	1 µL	500 µL
#1a DNA control tmtc2 #1	0 µL	500 µL	1 µL	500 µL
#2 Transfection tmtc2 #2	20 µL	480 µL	1 µL	500 µL
#2a DNA control tmtc2 #2	0 µL	500 µL	1 µL	500 µL
#3 Transfection tmtc2 #3	20 µL	480 µL	1 µL	500 µL
#3a DNA control tmtc2 #3	0 µL	500 µL	1 µL	500 µL
#4 Transfection tmtc2 #4	20 µL	480 µL	1 µL	500 µL
#4a DNA control tmtc2 #4	0 µL	500 µL	1 µL	500 µL
#5 Transfection tmtc3 #1	20 µL	480 µL	1 µL	500 µL
#5a DNA control tmtc3 #1	0 µL	500 µL	1 µL	500 µL
#6 Transfection tmtc3 #2	20 µL	480 µL	1 µL	500 µL
#6a DNA control tmtc3 #2	0 µL	500 µL	1 µL	500 µL
#7 Transfection tmtc3 #3	20 µL	480 µL	1 µL	500 µL
#7a DNA control tmtc3 #3	0 µL	500 µL	1 µL	500 µL
#8 Transfection tmtc3 #4	20 µL	480 µL	1 µL	500 µL
#8a DNA control tmtc3 #4	0 µL	500 µL	1 µL	500 µL
#9 Transfection sH control 1	20 µL	480 µL	1 µL	500 µL
#9a Lipofectamine only	20 µL	500 µL	0 µL	500 µL
#10 Transfection sH control 2	20 µL	480 µL	1 µL	500 µL
#10a Control – no treatment	0 µL	500 µL	0 µL	500 µL

The transfection mixture was added to each dish and dishes were rocked gently to ensure even coverage. The dishes were incubated at 37 ºC under 5% CO_2 overnight. The

next day, cells were washed with PBS, trypsinized using 0.5 mL trypsin-EDTA as above and each transfection was transferred into a fresh 25 cm^2 culture flask, for a total of 20 flasks. Fresh D-MEMs, 4.5 mL, was added to each flask and the cells were incubated for 24 h at 37 ºC under 5% CO_2.

The following day, 1 µg/mL puromycin (pur) was added to each 25 cm^2 flask, and the cells were incubated at 37 ºC under 5% CO_2. After 2-3 days, cells in the controls (DNA only (#a), LF only, or no treatment) were observed to have died. Transfection reactions (1-10) also experienced some cellular death, as noted by floating cells in the media. All samples were grown in D-MEMs supplemented with 1 µg/mL pur, and incubated at 37 ºC under 5% CO_2. Cells were monitored by light microscopy and trypan blue viability assays, and transferred into new flasks when necessary.

5.2.21 POTENTIAL C-MANNOSYLTRANSFERASE TMTC2 AND TMTC3 SILENCING IN HEK-293T

When cellular growth was observed in flasks, media was removed from transfected cells and the cells were washed with PBS. Trypsin-EDTA (0.5 mL) was introduced to each culture flask, and the flasks were incubated for 5 min at 37 ºC under 5% CO_2. After incubation, cell suspensions were triturated to disperse clumps and remove any remaining adherent cells from the flasks. Cell suspensions were transferred into 5 75 cm^2 flasks and 14.5 mL of D-MEMs was added. After overnight incubation at 37 ºC under 5% CO_2, media was replaced with 15 mL of fresh D-MEMs supplemented with 1 µg/mL pur. Cells were grown at 37 ºC under 5% CO_2 in D-MEMs supplemented with 1 µg/mL pur until cells reached 20-40% confluence. At this point, cells were again washed with PBS and trypsinized for transfer to 3 185 cm^2 flasks. Cells were allowed to grow in 25 mL of D-MEMs supplemented with 1 µg/mL pur at 37 ºC with 5% CO_2.

5.3 RESULTS AND DISCUSSION

We have proposed that the TMTC2 and TMTC3 proteins are the *H. sapiens C*-mannosyltransferases. To test this hypothesis a biological system with an endogenous *C*-mannosylation pathway that did not recognize the WXXW motif was used for heterologous expression of the *tmtc2* and *tmtc3* genes. A gene encoding a *C*-mannosylation targeted protein was coexpressed in the same host. The gene products were isolated and subjected to an assay to detect *C*-mannosylation. Previously, these genes were expressed in *S. cerevisiae*. However, difficulties were encountered in both obtaining microsomes and co-expression of the chosen protein target gene in yeast. Insect cells offer a convenient platform for *in vivo* assay through co-expression of target genes as well as the ready availability of insect cell microsomes for use in assays *in vitro*. Therefore, we generated bacmid constructs containing *tmtc2*, *tmtc3* and a dual expression construct for *tmtc2* and *tmtc3* together for gene expression in insect cells.

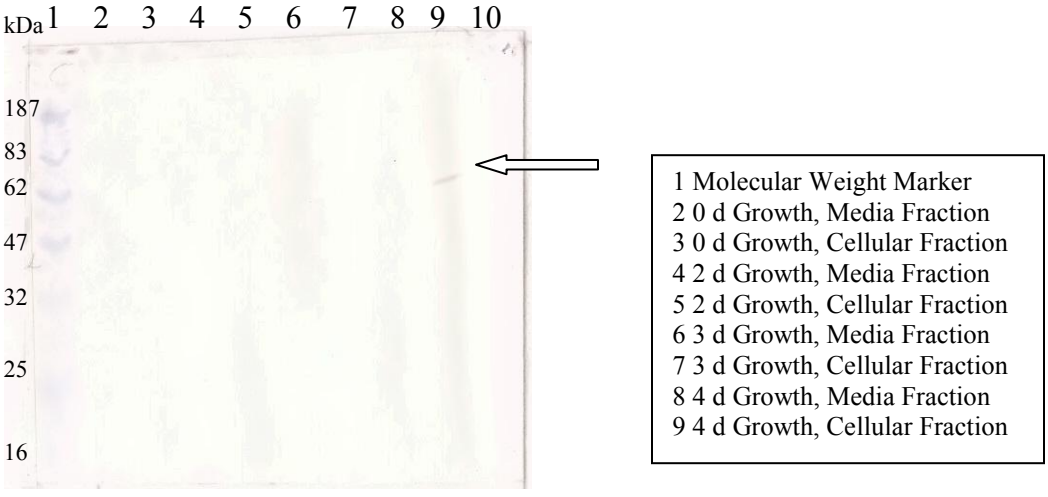
5.3.1 EXPRESSION OF TMTC2 AND TMTC3 IN SPODOPTERA FRUGIPERDA (SF9) CELLS

Sf9 insect cells have been previously shown to lack *C*-mannosylating activity against the canonical WXXW sequence as found in RNase2.[228] Therefore, to investigate whether *tmtc2* and *tmtc3* are *C*-mannosyltransferases, they were co-expressed with the target protein, IL-12b, which is known to be *C*-mannosylated *in vivo*. Ideally, the TMTC2 and TMTC3 proteins generated through baculovirus infection would be targeted to the endoplasmic reticulum, and thus have access to the Dol-P-Man substrate. The *il-12b* gene, in a plasmid expression construct with an insect secretion signal, would also be translocated into the ER of the insect cells. The IL-12b protein would thus have the

opportunity to be modified *in vivo* while passing through the secretory pathway prior to secretion into the culture media. The IL-12b protein, with a *C*-terminal His_6 tag could then be harvested from the media, using Ni-NTA resin, and analyzed by MALDI-MS for *C*-mannosylation.

Both media and cellular fractions were analyzed for TMTC2 and IL-12b proteins to ensure that proteins were retained in their predicted locations: TMTC2 within cells and IL-12b secreted into the media. The TMTC2 protein was found to be enriched in the cellular membrane fractions as determined according to the method of Schutzbach, which employs the addition of increasing concentrations of detergents SDS and Nonidet P40. [124] Expression of the gene *tmtc2* alone was followed using SDS-PAGE and Western Blot over the course of 4 d. This time course study demonstrated that TMTC2 protein was produced 4 d after infection, as shown in the Western Blot in 5.1.

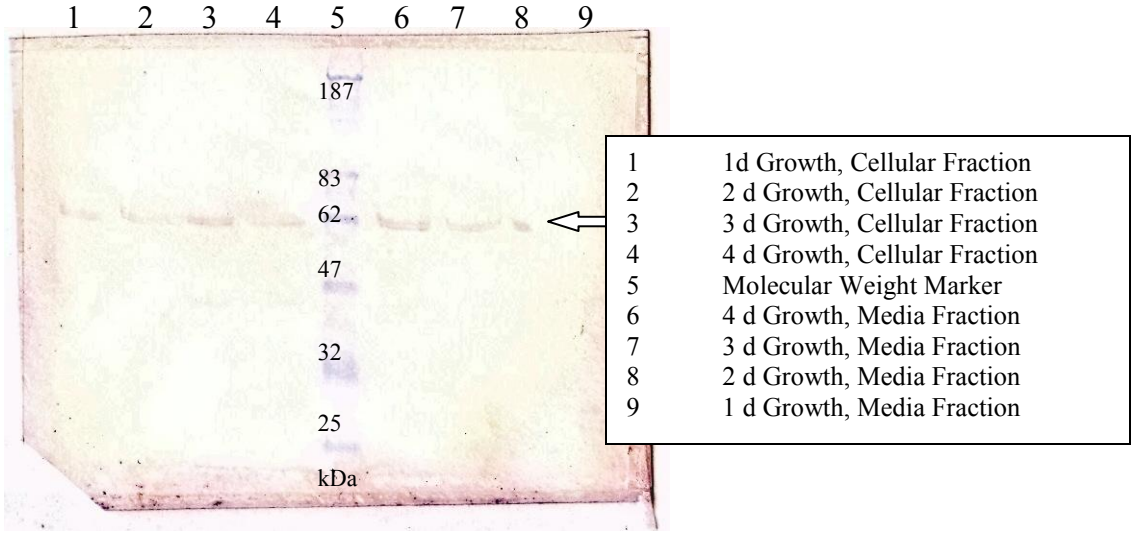
5.1 TMTC2 Production in Sf9 Cells (Time Course Assay)



We did not expect such a delay in protein production, as many other proteins generated using the baculovirus system are present within 2-3 d post-infection. [367]

Expression of the *il-12b* gene was also examined. For *in vivo* assays to be effective, it is imperative that this protein passes through the secretory pathway, and thus be available for harvest from the media. To assess IL-12b production, media samples from a 4 d time course assay were concentrated and subjected to Ni-NTA batch extraction to isolate the His_6-tagged protein. It should be noted that culture conditions for IL-12b expression deviated from the norm as L-glutamate strips Ni^{2+} from Ni-NTA beads, thus the culture conditions for TMTC2 and IL-12b generation in these time course experiments were not identical. The crude cell fractions and elution fractions obtained from Ni-NTA batch phase extraction of the media were examined using SDS-PAGE and Western Blot.

5.2 IL-12b Production in SF9 Cells (Time Course Assay)



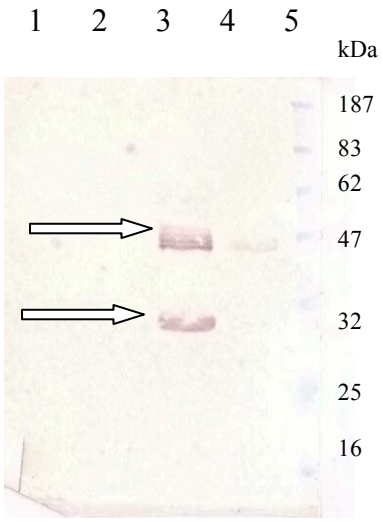
The IL-12b protein was found in both the cell and media fractions. As shown in Figure 5.2, IL-12b appears in the cellular fraction within 1 d and remains throughout the 4 d time course. However, the protein only appears in the media beginning at 2d of growth.

As the cells used for this experiment are stably transfected with an expression vector carrying the *il-12b* gene, we expect that the IL-12b protein should be found in all cellular samples from 0-4 d, as was observed. Our concern, however, is the difference in timing of expression of the *tmtc2* gene versus the *il-12b* gene. If the *il-12b* gene is expressed and the protein is secreted faster than the TMTC2 protein is produced, then the majority of our IL-12b protein will pass through the secretory pathway before *C*-mannosylation can take place.

To ensure that the protein isolated from the Sf9 cells was in fact IL-12b, we grew transfected cells in suspension culture under the same conditions as above for 2 d in media supplemented with BSA for protein stability. The media was harvested and the IL-12b protein was extracted from the media using Ni-NTA resin as described above. Fractions from the media, Ni-NTA washes, and elutions were examined by SDS-PAGE and Western Blot. The Western Blot showed two very clear bands in the elution fraction as shown in Figure 5.3. This observation differs from results described above, and we speculated that the additional band was BSA or a component of BSA. Both bands from the SDS-PAGE gel were isolated and subjected to in-gel digestion and mass spectrometry analysis to confirm the identity of the IL-12b protein. MS results did identify the new protein band as a bovine protein. It is most likely a component of the BSA mixture added to the media for protein stability. Unfortunately, MS results for the putative IL-12b protein band revealed that the band contained a complex protein mixture and thus was not

suitable for protein identification at this level of purity. For these reasons, we were unable to confirm the production of the IL-12b protein by the Sf9 cells.

5.3 IL-12b His-Tag Purification from Sf9 Media (2 d Production)

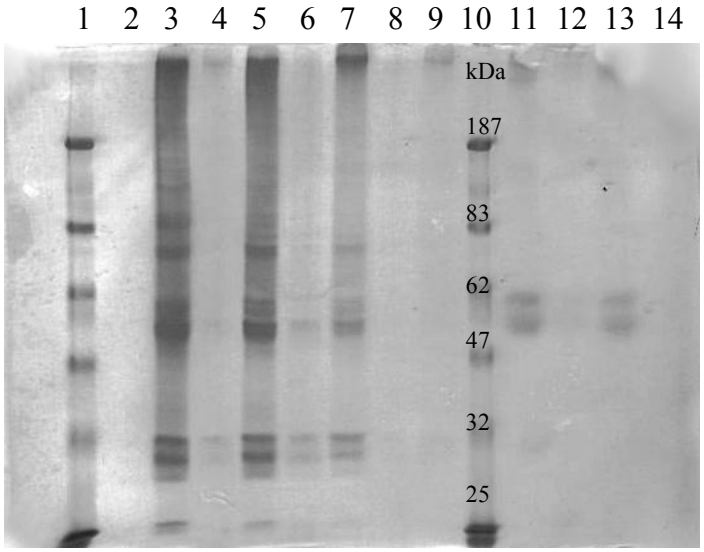


1	Ni-NTA Flow Through
2	Ni-NTA Wash
3	Ni-NTA Elution
4	2 d Media Fraction
5	Molecular Weight Marker

Due to the uncertainty regarding the expression levels and timing of both genes in Sf9 cells. We decided to co-express genes for the potential *C*-mannosyltransferase TMTC2 and the target protein IL-12b in insect cells. We attempted coexpression by using the TMTC2-baculoviral stock to infect cells that were already stably transfected with the pMIB/IL-12b plasmid. Cells were grown in suspension culture under the same conditions as described above. After time course assay over 4 d, both media and cell samples were assessed for the production of TMTC2 and IL-12b. The media was concentrated and subjected to protein enrichment using Ni-NTA beads to isolate IL-12b, while the membrane fractions containing our TMTC2 protein of interest were enriched from the cells using the method described previously.[124] The elution fractions from the Ni-NTA enrichment and C/D fractions from the membrane enrichment were analyzed by SDS-PAGE and Western Blot. A 1 L culture of Sf9 cells

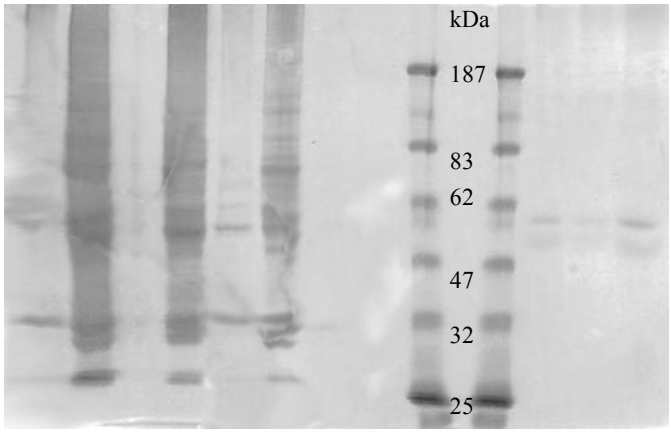
without the pMIB plasmid or TMTC2 viral stocks was grown in parallel as a negative control.

5.4 Sf9 Negative Control (cells only, Time Course)



1	Molecular Weight Marker
2	96 h Cellular Fraction C
3	96 h Cellular Fraction D
4	72 h Cellular Fraction C
5	72 h Cellular Fraction D
6	48 h Cellular Fraction D
7	48 h Cellular Fraction C
8	
9	
10	Molecular Weight Marker
11	96 h Media Fraction
12	72 h Media Fraction
13	48 h Media Fraction
14	0 h Media Fraction

5.5 Sf9 + TMTC2/IL-12b (Time Course)



1	96 h Cellular Fraction C
2	96 h Cellular Fraction D
3	72 h Cellular Fraction C
4	72 h Cellular Fraction D
5	48 h Cellular Fraction D
6	48 h Cellular Fraction C
7	
8	
9	Molecular Weight Marker
10	
11	Molecular Weight Marker
12	96 h Media Fraction
13	72 h Media Fraction
14	48 h Media Fraction

From the SDS-PAGE, we were not able to observe the presence of either protein, and the Western Blots confirm these results. We hypothesized that a potential problem with protein production is the cell line used. Sf9 was chosen because it does not exhibit inherent *C*-mannosylation on WXXW motifs. However, this cell line is a very poor choice for gene expression. It was chosen because it is also well suited for baculovirus expression and viral stock generation.[367] One common choice for secreted protein production is *Trichoplusia ni*, or High Five cells, however we did not choose to work in these cells as *C*-mannosylation was previously reported in this cell line.[368] For our purposes, we believed it would be better to confine ourselves to *S. frugiperda* cell lines, which have not been shown to exhibit *C*-mannosylation of the canonical *C*-mannosylation motif (WXXW). It was reported that Sf9 cells have intrinsic *C*-mannosylation machinery, although *C*-mannosylation events are limited to modification of thrombospondin repeat regions. The response of High Five cells to the canonical WXXW signal sequence is unknown, and therefore we could not ensure that any *C*-mannosylation that may be observed on our target protein is due to the heterologous expression of *tmtc2* or *tmtc3*.

5.3.2 C-MANNOSYLTRANSFERASE TMTC2 AND TMTC3 EXPRESSION IN SPODOPTERA FRUGIPERDA (SF21) CELLS

As previously stated, Sf cells are not ideal for producing secreted proteins; therefore, we decided instead to focus on an *in vitro* approach. This entailed expression of *tmtc2*, *tmtc3* and the dual expression of *tmtc2/3* in insect cells, followed by purification of endoplasmic reticulum as microsomes. To alleviate the difficulties in protein production, the Sf21 cell line was chosen. It is unknown whether this cell line exhibits *C*-mannosylating activity; as it has never been tested in this capacity. Nevertheless, the

protein production capacity of this cell line is significantly greater than that of Sf9 cells. Therefore, to ensure the validity of any observed *C*-mannosylation in microsomes, we also purified microsomes from cells with no additional expression constructs to serve as controls for background activity.

Sf21 cells were cultured as adherent monolayers and infected with P2 baculovirus stocks for the expression of *tmtc2*, *tmtc3* and the dual expression of *tmtc2/3*. Cells were harvested at various time points during growth to examine protein production and crude cellular fractions by SDS-PAGE without purification. Small (10 µL) aliquots of cells were removed and subjected to SDS-PAGE and Western Blot. From these samples, it was determined that 2-3 d growth yielded the best production of the proteins.

5.3.3 Purification of Microsomes Containing the Potential C-mannosyltransferases TMTC2 and TMTC3 from Spodoptera frugiperda (Sf21) Cells

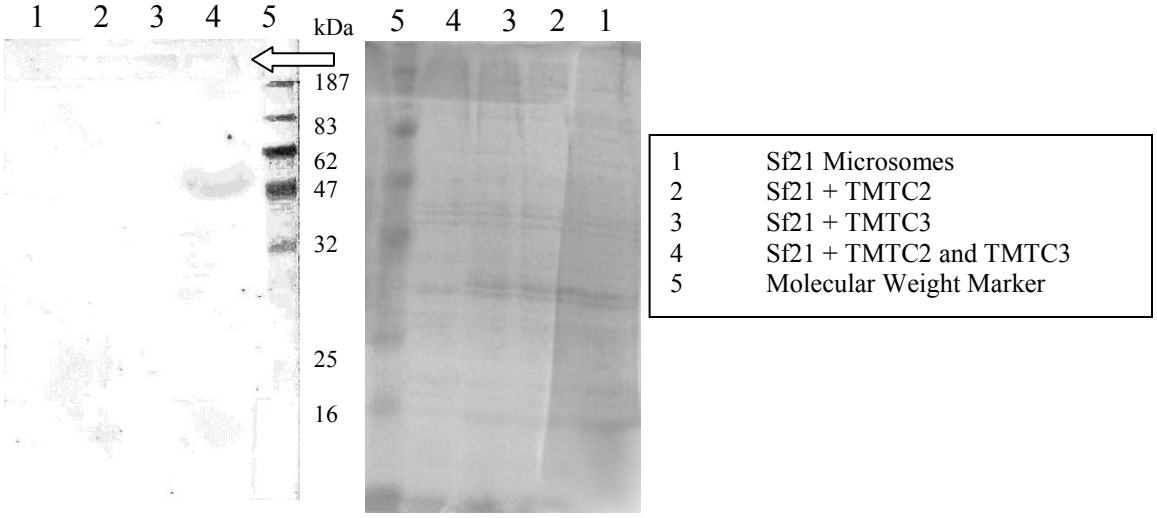
The TMTC2 and TMTC3 proteins are multipass transmembrane proteins, and if they exhibit CMT activity, they will work on a membrane bound substrate, Dol-P-Man. Purification of these proteins presents several challenges. Proper orientation of protein domains may be affected by purification. Purification with detergents may perturb the folding and structure of protein domains not contained within the membrane. Once the protein is isolated from the membrane fraction of the cells, it is also removed from the membrane fractions that contain its natural substrate, thus limiting the protein's access to substrate. Furthermore, the substrate Dol-P-Man is no longer commercially available and must be synthesized from dolichol using both chemical and biological methods or directly extracted from microsomes and added into the reaction. These difficulties may be

overcome by using microsomes, because both the Dol-P-Man substrate and the TMTC2/3 proteins are present and in the proper orientation for activity.

To generate microsomes containing TMTC2, TMTC3 and TMTC2/3, cells were grown as described in the previous section. Different methodologies were utilized to purify microsomes, as described previously. Generally, cells were resuspended in a lysis buffer and manually disrupted (homogenized) to break the cell membranes. Low-speed centrifugation was employed to remove cellular debris and large organelles (nuclei and mitochondria). The supernatant from the low-speed centrifugation was subjected to high-speed centrifugation, which removed soluble proteins from the cell lysate. The pellet from this ultracentrifugation step contained the endoplasmic reticulum and could be further purified through additional ultracentrifugation employing a sucrose gradient step to separate heavier rough endoplasmic reticulum from lighter smooth ER fractions, although a crude microsomal fraction was sufficient for our purposes.

The microsomal pellets containing rough ER or total ER fractions were resuspended in buffer (usually HEPES based with multiple protease inhibitors) and the total protein concentration was determined by Bradford assay. Microsomal fractions had average protein concentrations at approximately 2 mg/μL, which varied slightly depending on the purification method used. The microsome samples were divided into aliquots, flash frozen, and stored at –80 ºC until use. Due to the complexity of the mixture, when SDS-PAGE was attempted on these samples, the protein bands were unresolvable. The protein samples were too concentrated to visualize any discrete bands, and dilution of these samples until discrete bands were seen yielded very poor Western blot results, although we could see very faint bands at the top of the membrane.

5.6 Sf21 Microsomes Containing TMTC2 and TMTC3 Coomassie Gel and Western Blot



5.3.4 INSECT CELL MICROSOMAL ASSAYS TO DETECT C-MANNOSYLTRANSFERASE ACTIVITY

By taking the ER portion of the secretory pathway, we can then add any designed peptide substrate to the assay, thus overcoming the issue of targeted protein production. Thus an advantage of this assay design is that we are no longer limited by naturally occurring substrates. Potentially, any synthetic substrate with the WXXW motif could be used in this assay however, we chose the protein substrate derived from RNAse2 as it had been previously shown to be an accepted peptide substrate for *C*-mannosylation.[126] The microsomes from the previous section, which contained the proteins believed to have *C*-mannosylating activity, were incubated overnight with the synthetic peptide substrate, 0.2% Triton X-100 to solubilize microsomes, and a protease inhibitor cocktail in HKA

(Hepes-potassium acetate) buffer. Assays were incubated both at room temperature and at 37 ºC.

After incubation the reaction solution was filtered to remove proteins above 10 kDa and an attempt to isolate the peptide substrates by HPLC was made. Using a linear gradient of water/0.1% TFA and 90% acetonitrile/ 0.1% TFA we were able to establish a retention time for the unmodified substrate peptide ~ 37 min. The peptide substrate used has two tryptophan residues and absorbances at 280 and 214 nm. Therefore, each assay sample was divided and run twice on the HPLC, collecting peaks displaying absorbance at both wavelengths. These samples were concentrated and submitted for MS analysis. The desired peptide could not be identified in the microsome assay samples submitted. The MS results show a mixture of peptides, none matching the expected or previously found molecular weight (m/z 1798) of the peptide. It appears that the purification method used was not suitable for the sample.

To overcome the purification issues, a biotin label was introduced at the lysine residue generating the peptide: N-AC-K^{biotin}PPQFAWAQWFE-NH_2. Introduction of this label improved our ability to recover the peptide from the complex microsomal mixture. The microsomal experiment was repeated, substituting the biotinylated peptide for unlabelled substrate, and the peptide was purified using streptavidin beads.

The MS analysis of isolated biotinylated peptide show some differences between the control samples and the experimental samples, however the expected increase of 162 Da was not observed in any samples. We observed an addition of 114 Da in every experimental sample (TMTC2, TMTC3, and TMTC2/TMTC3). We did not see the addition in the control sample (no TMTC2/TMTC3). It is possible that the harsh elution conditions (0.1% TFA) cause a fragmentation of the mannose pyranose ring, thus altering

the observed molecular weight of the modified peptide. These are promising results, but certainly not definitive enough to conclude that *C*-mannosylation is occurring.

It is also possible that the TMTC2 and TMTC3 proteins are not functional in insect cells due to a number of factors. First, the proteins may not be retained in the ER as hoped. In this case, the protein activity would not be observed using microsomal assays. To solve this problem, complete membrane fractions could be isolated from insect cells and applied to the assays. Complete membrane fractions would include the ER as well as the plasma membrane, thus providing both protein and Dol-P-Man substrate in the assay. Second, it is possible that the conditions in the insect cells do not mirror those found in human cells closely enough to facilitate *C*-mannosyl transfer. This could result from protein misfolding or key residues at improper protonation state, as the pH of the insect cell is 6.1-6.4, which is ten-fold more acidic than that of a human cell. Despite these possibilities, recovery of intact peptide from the control reactions demonstrated that the peptide substrate itself did not undergo cleavage or other modification under the assay conditions.

5.3.5 SILENCING PUTATIVE C-MANNOSYLTRANSFERASES

Another option that was explored was silencing those genes in human cells that are candidates for being *C*-mannosyltransferases. The expression of the *sGP* gene from *Zaire ebolavirus* (Zebov) was attempted, using human HEK-293T cells. Purification of this protein from mammalian cell culture would provide a positive control for *C*-mannosylation. This protein could be digested, and the peptides separated to isolate the *C*-mannosylated peptide. Silencing constructs could then be introduced to the HEK-293T cells and the *sGP* expressed again and examined for *C*-mannosylation. A lack of *C*-

mannosylation on the protein purified from the silenced cell lines would be strong evidence that the proteins TMTC2 and TMTC3 were are very likely responsible for *C*-mannosylation.

The *sGP* gene was amplified by PCR and ligated into plasmids for expression in mammalian, insect, yeast and bacterial cells. Attempts to express the gene in yeast and insect cells were unsuccessful. To our knowledge, no previous attempts had been made to express *sGP* in yeast cells. Surprisingly, *S. cerevisiae* died when transformed with the plasmid containing *sGP*. Despite multiple attempts at transformation both in liquid culture and on plates, no growth of treated yeast cells was observed. Likewise, no sGP protein could be obtained from Sf21 insect cells transfected with the *sGP* gene.

When expressed in *E. coli*, the protein was found to be an aggregate, most likely due to the lack of *N*-glycosylation and thus in proper protein folding in *E. coli*. Properly folded protein may not be necessarily required for our assays, as the goal is to assay for *C*-mannosylation rather than protein activity. Thus, we persevered and attempted to purify the His_6 tagged sGP protein using Ni-NTA resin under both native and denaturing conditions. Both attempts failed, and sGP eluted in the wash fractions. The protein appeared to have no affinity for the Ni-NTA beads, although hybridization with α-His antibodies clearly demonstrated the presence of the tag by Western Blot. In cases such as this, often the polyhistidine tag is buried and inaccessible. However, if this is the reason the sGP protein should have bound to the resin under the harsh denaturing conditions employed in the second attempt. We reasoned that the protein does not bind to the resin due to its highly positively charged nature. At pH 7, sGP carries a charge of +12, which is significantly different from most soluble proteins. The isoelectric point of the protein is approximately 9.4, thus a Tris buffer with a pH of 8 at room temperature (or 8.5 at 4 ºC) may potentially facilitate salting out of our protein as we encroach on the isoelectric point

of sGP. A viral extraction of adenovirus producing sGP was noted to have a bright yellow hue to the solution (culture media), indicating a pH far below the neutral 7. A small sample of the viral preparation was then obtained and dialyzed against pH 7 buffer until the sample turned pink. At this point, protein precipitate was observed in the solution. This observation suggested that sGP is not stable at higher pH. Thus, buffer conditions were adjusted accordingly to lower pH.

Next, the His_6-tag was circumvented, and purification using size exclusion, cation and anion exchange columns was attempted instead. In all cases, sGP either did not bind the column or precipitated during elution. A final method using low pH buffer and ammonium sulfate precipitation, followed by DEAE anion exchange chromatography was also attempted. While this approach was successful by SDS-PAGE and Western Blot, MS was unable to identify the peptide of interest, VNPEIDTTIGE***W****AFW*ETK(K) (m/z 2136, 2264).

Attempts to express the sGP protein in mammalian cells were much more successful. Confluent cells transfected with the pEF/*myc*His-sGP construct were grown in D-MEMt supplemented with 10 μg/mL bsd under standard conditions, and time course assays were performed to determine the optimum time to harvest media for protein purification. SDS-PAGE and Western Blot results indicated that a 4 d incubation yielded the highest level of protein production. It should be noted that the 3 and 4 d sample media were quite yellow, indicating the pH of the culture media had become much lower than the optimal (slightly) basic pH for normal cell growth.

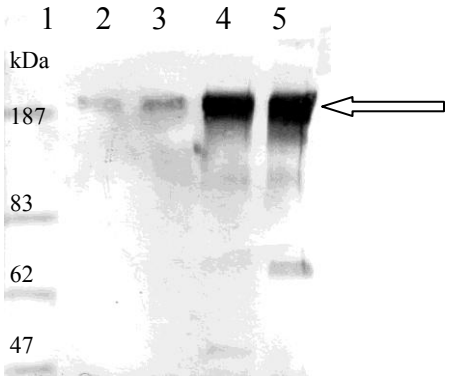
The media was concentrated and subjected to a variety of purification methods. These include cation exchange, anion exchange, size exclusion, and lectin chromatography. In contrast to expression in *E. coli*, we believed that the protein generated by mammalian cells should be more stable in solution due to its glycosylation

state, and thus be properly folded. Since the protein at pH 7 should be positively charged, we attempted cation exchange chromatography using Sepharose-Q and CM-Cellulose resins. The protein again precipitated in both columns under high salt concentrations, thus we could not elute the protein from either of the cation exchange columns.

Lectin chromatography had previously been used to purify sGP from culture media, thus media concentrate was subjected to affinity chromatography using concanavalin A lectin, which binds to the terminal mannose residues of glycoproteins.[230, 362] This procedure failed as the protein did not bind to the ConA beads, but was eluted from the column under wash conditions. This result implies that of the 5 *N*-glycans present on the protein, there are no terminal mannose residues on the sGP protein when purified from HEK-293T cells. It is possible that this discrepancy may arise from the use of different cell lines. Published procedures from Barrientos et al. and Falzarno et al. used sGP protein produced by Vero cells for lectin affinity purification methods, while the current work employed HEK-293T cells.[230, 362]

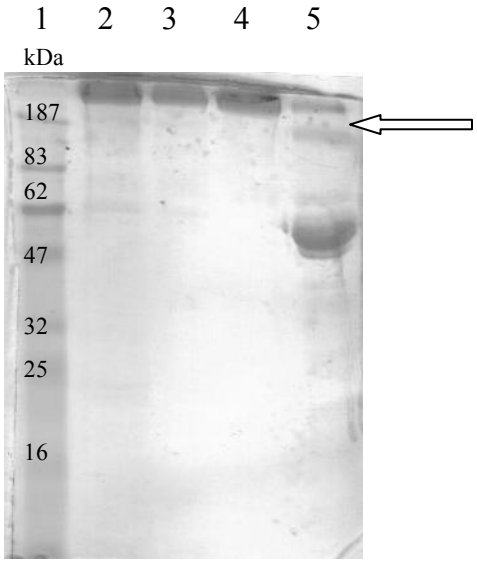
The next purification attempt involved anion exchange chromatography. Tris buffer was chosen for these attempts based on previous successful use of Tris-NaCl (TN buffer) in hemagluttinin (HA) affinity chromatography.[230] Unfortunately, the protein again precipitated. Changing the buffer to MES pH 5.5 yielded improved results, and the protein appeared in the first 2-4 "flow-through" fractions of the DEAE anion-exchange chromatography. This adequately separated the target sGP protein from the BSA in the culture media. Fractions containing the protein were subjected to size exclusion chromatography using an FPLC equipped with a Sephadex 200 column. MES (pH 5.5) buffers were used in the isocratic elution. Fractions judged to contain the protein were combined based on purity, flash frozen, and stored at −80 ºC (see Figure 5.8)

5.7 Western Blot of sGP DEAE Column Fractions 1-4



1	Molecular Weight Marker
2	DEAE Fraction 1
3	DEAE Fraction 2
4	DEAE Fraction 3
5	DEAE Fraction 4

5.8 sGP DEAE Column Fractions and Sephadex Purification



1	Molecular Weight Marker
2	Sephadex Fractions (Trial 1)
3	Sephadex Fractions (Trial 2)
4	Sephadex Fractions (Trial 3)
5	DEAE Fractions1-4

5.3.6 sGP IN-GEL DIGESTION AND ANALYSIS BY MASS SPECTROMETRY

This protein would serve as the positive control for future silencing experiments, in which the activities of TMTC2 and TMTC3 proteins will be knocked down. Thus it was imperative that not only the protein be identified positively, but the modified peptide be

detected using mass spectrometry. Accordingly, the sGP protein isolated from HEK-293T cells was subjected to in-gel tryptic digestion and mass spectrometry to confirm the identity of the protein. Unfortunately, the protein in question could not be identified after repeated rounds of MALDI-TOF and MALDI-TOF/TOF analysis. Most importantly, the potential *C*-mannosylated peptide in question: VNPEIDTTIGE***W****AFW*ETK(K) (m/z 2136, 2264) was not observed. Even in the previously published work, this peptide was barely detected by spectroscopy, quite close to background.[230] Undaunted, we attempted to purify the peptide from both in-gel digestions and solution digests using HPLC as described previously. These efforts were unsuccessful and MS results showed a mixture of peptides in the samples, but no peaks corresponding to the peptide of interest, either modified (m/z 2298, 2426) or unmodified (m/z 2136, 2264) were observed.

5.3.6.1 N-Azido Sialyl Labeling of sGP in HEK-293T

We next attempted to express the *sGP* gene in the presence of *N*-Aza-tetraacetyl-mannose (Ac_4ManNAz), such that glycosylated proteins would be fractionally labeled with *N*-azido-sialic acid. Previously, the Bertozzi group had shown that tetraacetylated azidomannose could be taken up by cells under normal culture conditions with low glucose. The cellular machinery could convert the azido sugar to SiaNaz which is the incorporated into complex glycans associated with many types of glycoproteins.[369] Through this expression method, the introduced azide moiety on the sialic acid residues could be coupled to Cu(I)-mediated alkyne-biotin conjugates and the resulting proteins could be isolated using streptavidin affinity chromatography. The protein of interest could be further purified from this mixture of proteins using methods already developed. sGP has 6 known glycosylation sites, five of which are believed to have one or more

terminal sialic acid moieties.[362] However, two attempts to grow adherent monolayers in the presence of 30 and 15 μM tetraacetyl azidomannose (D-MEMm) were unsuccessful. In both cases all cells in the flasks died within 4 - 7 d, as judged by morphology and non-adherence.

The cause of cell death is not clear, but may be a result of toxicity due to sGP. We had noted previously that monolayer cultures producing sGP were not as healthy as the control cells. Compared to control cells grown under the same culture conditions, cells expressing sGP divided much more slowly. For example, a flask of HEK-293T control cells was passaged (1:10) approximately every week. Cells transfected with sGP were only confluent enough to passage every two weeks at a 1:2 split. These cells died after approximately 10-15 passages. From this we concluded that the production of sGP protein causes cell death, as compared to control HEK-293T cells grown in parallel under identical conditions. Under labeling conditions, control HEK-293T cells grown in parallel (low glucose supplemented with 4Ac-ManNaz) were not as robust. Cells under low glucose conditions were passaged approximately every week at a 1:4 split, but still grew. Cells producing sGP under labeling conditions never survived long enough to split, as they died within 4 days. Even under adjusted conditions, these cells survived for less than two weeks before cell death. We propose that the addition of ManNaz to these cells causes a stress on the cellular system that, when combined with the apparent toxicity of sGP, becomes too great a burden on the cell.

5.3.7 SILENCING TMTC2 AND TMTC3 IN HEK-293T CELLS

Due to the problems in obtaining the sGP protein or modified peptide for use as a positive control for *in vivo* assays with HEK-293T cells, we concentrated our efforts on

designing a new microsomal assay for use with synthetic peptides. In these experiments we would silence *tmtc2* and *tmtc3* in HEK-293T cells using shRNA. Microsomes from normal HEK-293T cells and silenced cells could then be harvested and applied to the previously described *in vitro* microsomal assay that employs a synthetic peptide. If we could silence the genes, thus preventing production of the TMTC2 or TMTC3 proteins, microsomes harvested from these cells could be investigated for a deficiency in *C*-mannosylation.

Briefly, 4 plasmids, each expressing a different shRNA designed to target either *tmtc2* or *tmtc3* expression were transfected into HEK-293T cells as described previously. Cells transfected with both control vectors (empty vector and non-effective silencing vector) recovered quickly and were at a confluence (70%) that could be passaged within a week of transfection. In contrast, cells transfected with silencing shRNA showed poor growth and recovery even after one month despite regular media exchange. Experimental cells never surpassed 10 – 20% confluence except in the case of constructs 1 and 4 for *tmtc2* and 4 for *tmtc3* silencing, which only reached 40%. In all experimental cases, cell death was observed continuously during growth as monitored by microscopy. The viability of cells transfected with silencing shRNA were examined using the trypan blue assay and in all cases were found to be less than 5% viable.

Even if transfection rates were originally low, cells would have recovered within several weeks of the procedure. Our conclusion is that regardless of whether *tmtc2* and *tmtc3* are indeed the *C*-mannosyltransferase genes or not, inhibition of the expression of these genes appears to be deleterious to human cells. Our original intent was to isolate microsomes from these cells, however this was not possible as silencing seems to cause cell death. These results are of some interest, as the deletion or removal of important glycosyltransferases such as β-OGT and POMT1/2 also leads to cell death. [205, 213, 352, 354]

5.4 CONCLUSIONS AND FUTURE DIRECTIONS

Our efforts to produce the TMTC2 and TMTC3 proteins in insect cells were successful; however, assays to detect *C*-mannosylation require more development. Our current efforts are focused on work in insect cell lines and attempts to perfect the microsomal assays with TMTC2 and TMTC3 proteins.

Silencing of these genes in the mammalian cell line HEK-293T has demonstrated interesting results. Whether these proteins are *C*-mannosyltransferases or not, it appears that their presence is essential to cells under culture conditions. This result is quite informative as it suggests that these predicted glycosyltransferases of unknown function may be required for cellular viability and growth. The functions of these proteins may be more important now than ever, as we search for the protein sources of CDGs with unknown etiology.

Once a suitable assay is fully developed, further characterization based on the methods employed by Hofsteenge can be undertaken. [227] Target proteins could be assayed for *C*-mannosylation by digesting with a protease such as elastase, and resultant peptide fragments fractionated by reversed phase HPLC using a C-18 column (acetonitrile/TFA gradient).[126] The purified fragments can then be subjected to MS and/or Edman degradation to determine the exact position of the mannose residue in the peptide. ^{1}H-NMR and TOCSY-NMR studies of the subpeptides can also be performed to verify the presence of a mannose residue. [370] Another proof of the presence of the C-C bond is acid hydrolysis followed by derivatization of the mannose. Since this residue is not attached to the tryptophan in a "typical" glycosidic linkage, acid hydrolysis will not break this bond. This method has been employed in gas chromatography/mass spectrometry (GC/MS) studies of *C*-mannosylation in which three divergent steps of GC/MS were used

in tandem with chemical modification. The methodology involves first a mild acid hydrolysis, followed by an acid-catalyzed methanolysis under anhydrous conditions, and finally a classical hydrolysis for peptide bond cleavage.[371] Derivatives of tryptophan were found to be stable under acid hydrolysis, while mannose derivatives were also stabilized as per-heptafluorobutyrate derivatives, and can be easily identified in the (positive) EI mode of ionization.[371]

Once a *C*-mannosyltransferase is identified, high throughput assays could be carried out to identify novel protein substrates, including proteins in infectious agents and the human immune system. The gene(s) could be knocked out and cellular viability, as well as the appearance of novel CDGs could be assessed. If we are correct in our assessment of the TMTC2 proteins, *C*-mannosylation is essential for cellular viability in higher eukaryotes. The enzyme mechanism of the protein could be probed, which would provide valuable insight into general *C*-glycosyltransferase mechanisms. The introduction of the CMT protein to heterologous systems could provide a platform by which to obtain *C*-mannosylated therapeutic proteins with improved serum half-lives. Above all, studies could be undertaken to elucidate the role of this novel posttranslational protein modification, *C*-mannosylation.

Appendices

ALIGNMENT OF TMTC2 HOMOLOGS USING CLUSTALX

```
Human           ------------------------------------------------------------
Mouse           ------------------------------------------------------------
Rat             ------------------------------------------------------------
Cow             ------------------------------------------------------------
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Dog             ------------------------------------------------------------
Chicken         MGKGQKADGSLCYWPCVKKHAPCEIPSGCSSLTAAPVHCLSSAVTAEIGPAPELWCELAC
CElegans        ------------------------------------------------------------


Human           ------------------------------------------------------------
Mouse           ------------------------------------------------------------
Rat             ------------------------------------------------------------
Cow             ------------------------------------------------------------
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Dog             ------------------------------------------------------------
Chicken         STLRAVTALMESCWQVVSSVKAMPTPRCLELYERHPEIKVGRCLEIRRRQDREGRKELWL
CElegans        ------------------------------------------------------------


Human           ------------------------------------------------------------
Mouse           ------------------------------------------------------------
Rat             ------------------------------------------------------------
Cow             ------------------------------------------------------------
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Dog             ------------------------------------------------------------
Chicken         NAWWLLGMWDAEDPEPVHRAGGKLPSQSCRGSGNLALTPAELIEIRAAALPVKRMLLPYF
CElegans        ------------------------------------------------------------


Human           ------------------------------------------------------------
Mouse           ------------------------------------------------------------
Rat             ------------------------------------------------------------
Cow             ------------------------------------------------------------
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Dog             -------------------------------------------MSMSEVGGGPGIFSGS
Chicken         NSVQAELKRTALLVYPEKLSKMLNTFVKFPYQHTDTGNCSLGKYRGDTKLVTYLILTSEI
CElegans        ------------------------------------------------------------
```

```
Human           ------------------------------------------------------------
Mouse           ------------------------------------------------------------
Rat             ------------------------------------------------------------
Cow             ------------------------------------------------------------
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------------------------------
Dog             SAGPAAEELPGGMIAGPRRSPRTCQDGPRRSPHTCQDGLRRSPHTCQDGPRRSPHTCQDG
Chicken         DDFSVFPLTEAECNQWELSVARLQFQIYALRDQSFQQSFSGDVSLELSYLCNTESPFLIK
CElegans        ------------------------------------------------------MKKE


Human           ---------------MIAELVSSALGLALYLNTLSADFCYDDSRAIKTNQDLLPETPWT-
Mouse           ---------------MIAELVSSALGLALYLNTLSADFCYDDSRAIKTNQDLLPETPWT-
Rat             ---------------MIAELVSSALGLALYLNTLSADFCYDDSRAIKTNQDLLPETPWT-
Cow             ---------------MIAELVSSALGLALYLNTLSADFCYDDSRAIKTNQDLLPETPWT-
Macaque         ---------------MIAELVSSALGLALYLNTLSADFCYDDSRAIKTNQDLLPETPWT-
Xenopus         ---------------MIAELLSSALGLLLYLNTLGADFCYDDSRAIKTNQDLLPETPWN-
Dog             PRRSPHTCQDGLRRSPHTCQDGPRRSPPKEALCPGSEALWMHTRAIKTNQDLLPETPWT-
Chicken         ITTELQFTFSISFLNVYCLFEQMEKKGSKCSALLQVIPDCAMSRAIKTNQDLLPETPWT-
CElegans        KRKQKSVEPTILSHIPYQYLLVALFASIVYRITLNADFVYDDRPAILTNDDVLGRTPWRS
                                                          ** **:*:* .***

Human           HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAIGGLNPWSYHLVNVLLHAAVTGLFTS
Mouse           HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAIGGLNPWSYHLVNVLLHAAVTGLFTR
Rat             HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAIGGLNPWSYHLVNVLLHAAVTGLFTR
Cow             HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAIGGLNPWSYHLVNVLLHAAVTGLFTN
Macaque         HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAIGGLNPWSYHLVNVLLHAAVTGLFTS
Xenopus         HIFFNDFWGTLLTHSGSHKSYRPLCTLSFRLNYLFGGLDPWNYHLVNVLLHSAVTGLFTN
Dog             HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRLNHAVGGLDPWSYHLVNVLLHAAVTGVFTS
Chicken         HIFYNDFWGTLLTHSGSHKSYRPLCTLSFRINHAIGGMDPWGYHLVNILLHAAVTGLFTN
CElegans        LIVHNDFWGNPIGLQGSHKSYRPLITASFRLQFAVHGLKAELFHGVNLICHMINSMLVLK
                 *..*****. :  .********* * ***::. . *:..  :* **:: *   : :.

Human           FSKILLGDG-YWTFMAGLMFASHPIHTEAVAGIVGRADVGASLFFLLSLLCYIKHCSTRG
Mouse           FSKALLGDG-YWTFMAGLMFASHPIHTEAVAGIVGRADVGASLFFLLSLLCYIKHCSTRG
Rat             FSKALLGDG-YWTFMAGLMFASHPIHTEAVAGIVGRADVGASLFFLLSLLCYTKHCSTRG
Cow             FSKILLGDG-YWTFMAGLMFASHPIHTEAVAGIVGRADVGASFFFLLSLLCYMKHCSTRG
Macaque         FSKILLGDG-YWTFMAGLMFASHPIHTEAVAGIVGRADVGASLFFLLSLLCYIKHCSTRG
Xenopus         LCKALFGSG-CWTLIAGLLFASHPIHTEAVSGIVGRADVGSGLFFLLSLLCYMKHCSTRG
Dog             FSKTLLGDG-CWTFMAGLLFASHPIHTEAVAGIVGRADVGASLFFLLSLLCYIKHCSTRG
Chicken         FSRILFGDG-YWTLIAGLLFASHPIHTEAVAGIVGRADIGACLFFLLSLMCYVKHCSTRS
CElegans        LARQMRIMGNEFSLLSALIFACHPITSEAVCSIVGRADLLSTMLILLAITCHTSDPSTFR
                :.: :   *  ::::::.*:**.*** :***..******: : :::**:: *: .. **

Human           YSARTWGWFLGSGLCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKIKQILPTIYKRKNLS
Mouse           YSARTWGWFLGTGLCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKMKQILPTIYKRKNLS
Rat             YSARTWGWFLGTGLCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKMRQILPTIYKRKNLS
Cow             SSARTWGWFLGTGMCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKMKQILPTIYKRKNLS
Macaque         YSARTWGWFLGTGLCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKIKQILPTIYKRKNLS
Xenopus         YSLSSWCWILCAGFWAACSMLWKEQGVTVLAVSAVYDVFVFHKLKMNQIISVVFKEKNVS
Dog             SSARTWGWFLGTGLCAGCSMLWKEQGVTVLAVSAVYDVFVFHRLKVKQILPAICKRKNLS
Chicken         SSARTWGWILGAGLCAGCSMLWKEQGVTVLAISAVYDIFVVHRLKMHQIIPALYKRKNLN
CElegans        -----------TIVLSILAVTAKETGIILLPLITLYDVLFKT-------------SNTKQ
                           : . :  ::  ** *: :*.: ::**::.               :. .
```

```
Human           LFLSISLLIFWGSSLLGARLYWMGNKPPSFSNSDNPAADSDSLLTRTLTFFYLPTKNLWL
Mouse           LFLSISLLTFWGTCLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFLYLPTKNLWL
Rat             LFLSISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWL
Cow             LFLSISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWL
Macaque         LFLSISLLTFWGSSLLGARLYWMGNKPPSFSNSDNPAADSDSLLTRTLTFFYLPTKNLWL
Xenopus         FFFSVGLLFAWGVILLGARFYWMGNTPPSFSNSDNPAADCEVLLTRTLTFFYLPTKNLWL
Dog             LFLSISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWL
Chicken         LFFSIGLLTSWGTALLGVRLYWMGNKPPSFSNSDNPAADSDSFLTRMLTFLYLPTKNLWL
CElegans        FRRDVTAYLISLLALCYLRLSINNFQSPKFSKNDNPIAHEPNSLTRALTFMYLPVFHLNL
                :  .:         *   *:   .  .*.**:.*** *.    *:* ***:***. :* *

Human           LLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYTGLLLLAYYGLKSPSVDRECNGKTVT
Mouse           LLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYSGLLLLAYCGLKNPSLEGECNGKALT
Rat             LLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYSGLLLLAYCGLKSPSLEGECNGKAVT
Cow             LLCPDTLSFDWSMDAVPLLKTISDWRNLHTVAFYTGLLLLAYFGLKSPSIERECNGKFVT
Macaque         LLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYTGLLLLAYYGLKSPSVDRECNGKTVT
Xenopus         LFCPDTLSFDWSMDAVPLIKTITDWRNIHTVAFYILLILLAYSSLKGSAIKRDCNGKVFM
Dog             LLCPDTLSFDWSMDAVPLLRTVCDWRNLHTVAFYTGLLLLAYYGLKSPGVERECNGKVVT
Chicken         LFCPDTLSFDWSMDAVPLLKAVSDWRNLHTVAFYAGLFLLAYFSLKGSSNERECNGKTVM
CElegans        IVFPKTLSYDWSMDAIPKVESLIDSRIILTFIVIG--------------VGEHVLLKLLL
                :. *.***:******:* :.:: * * : *. .                  .   * .

Human           NGKQNANGHSCLSDVEYQNSETKSSFASKVENGIKNDVSQRTQLPSTENIVVLSLSLLII
Mouse           NGKQNANGHSCHSDVEYRNSEMKPSFASKVENGIKNCVPQRTQLPSTENIVILSLSLLII
Rat             NGKQNANGHSCYSDVEYRNSEIKPSFASKVENGIKSCVSQRPQLPSTENIVILSLSLLII
Cow             NGKQNANGHSCHSEVEYGNSEIKPSFASKVENGIRNNASQRTQLPSTENIVVLSLSLLII
Macaque         NGKQNANGHSCLSDVEYQNSETKSSFASKVENGINNDVSQRTQLPSTENIVVLSLSLLII
Xenopus         NGKQNTNGHSCQSDLEHKNAEQNPVIASKLENGVKHHNSHEMQLPSTENIVVLALSLLIV
Dog             NGKQNANGHSCHSDGEYRNSESKPSLAAKAENGIKNNAAQRSPLPSTENIVVLSLSLLII
Chicken         NGKQNANGHSCHSEMEYKTLEGKSSFASKEENGIKKHEILKLQLPSTENIVILSLSLLIV
CElegans        GLRNSENSHENRS-----------------------------------LLFLFALFTT
                . ::. *.*.  *                                    ::: ::*:

Human           PFVPATNLFFYVGFVIAERVLYIPSMGFCLLITVGARALYVKVQKRFLKSLIFYATATLI
Mouse           PFIPATNLFFYVGFVIAERVLYIPSMGFCLLITVGARALYVKVQKRFLKSLVFYATATLI
Rat             PFIPATNLFFYVGFVIAERVLYIPSMGFCLLITVGARALYVKVQKRFLKSLVFYATATLI
Cow             PFVPATNLFFYVGFVIAERVLYIPSMGFCLLITVGARALYVKVQKRFLKSLIFYATAALI
Macaque         PFVPATNLFFYVGFVIAERVLYIPSMGFCLLITVGARALYVKVQKRFLKSLIFYATATLI
Xenopus         PFVPASNLFFYVGFVIAERVLYIPSMGFCLLVTVGARALYIKAQKNILKNLLFYATAALI
Dog             PFVPATNLFFYVGFVIAERVLYIPSMGFCLLVTVGARALYVKVQKRFLKSLIFYATATLI
Chicken         PFIPATNLFFYVGFVIAERVLYVPSMGFCLLVTVGVRALYVKAQKRFLKNLVFCSTAALI
CElegans        PHILSSNLLTHVGFVAAERILYLNTVAYCILAAFLAEMCSKRSSP----KSVLPLYILLL
                *.: ::**: :**** ***:**: ::.:*:* :. ..    : .     . ::     *:

Human           VFYGLKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAESAYRNALYYR
Mouse           VFYGVKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAESAYRNALFYR
Rat             VFYGLKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAEIAYRNALYYR
Cow             VFYGLKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAESAYRNALYYR
Macaque         VFYGLKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAESAYRNALYYR
Xenopus         VFYGLKTVVRNGDWKNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKIDEAENAYRNALYYR
Dog             VFYGLKTAIRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAERAYRNALYYR
Chicken         VFYGLKTVVRNGDWQNEEMLYRSGIKVNPAKAWGNLGNVLKSQSKISEAESAYRNALYYR
CElegans        SLFTIRTMQRVDDWKTEESLFKSALEVNPTKANMNLGYVYTTQKKYELAKYHYRQALKRQ
                 :: ::*  * .**:.** ***:*.::***:**  *** * .:*.* . *:  **:**  :
```

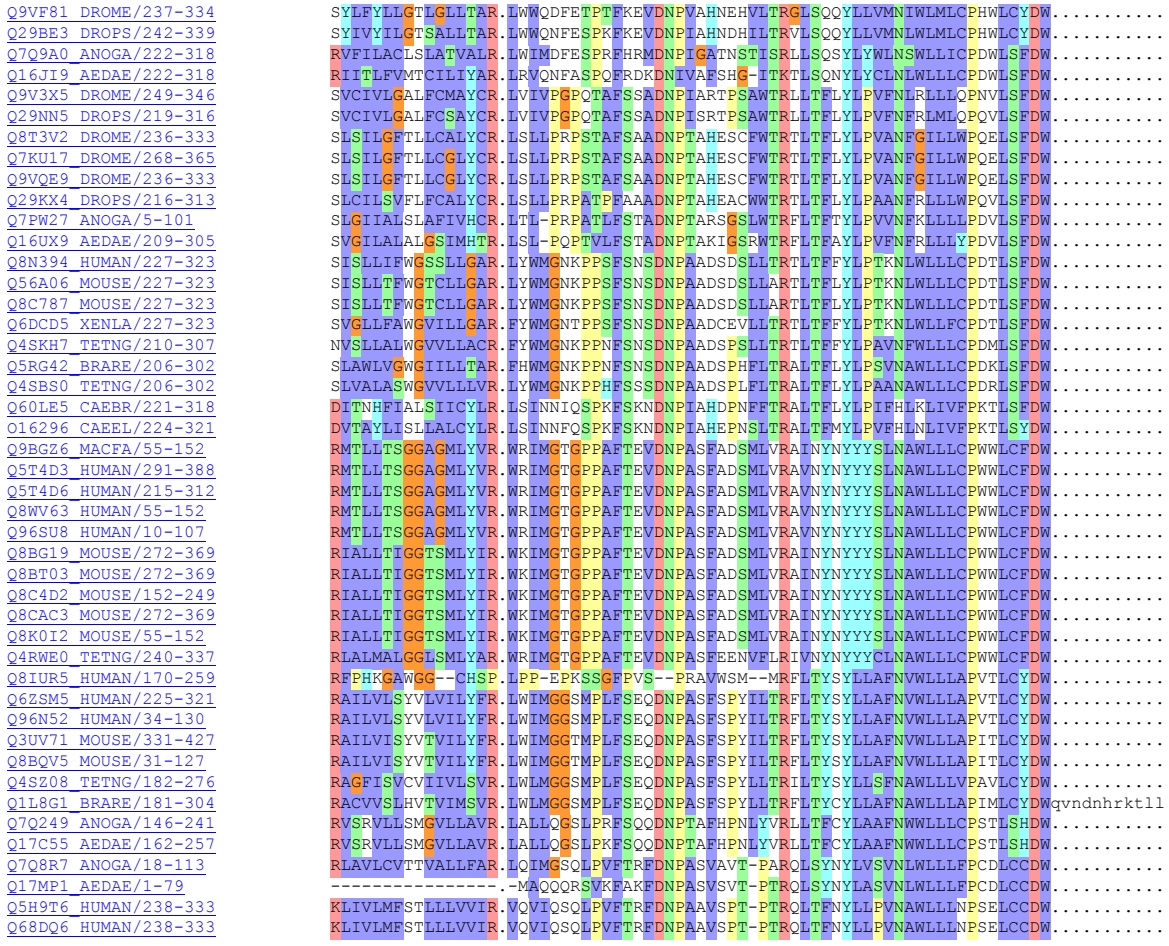
```
Human           SNMADMLYNLGLLLQENSR-FAEALHYYKLAIGSRPTLASAYLNTGIILMNQGRTEEARR
Mouse           SNMADMLYNLGLLLQENSR-FAEALHYYKLAIGSRPTLASAYLNTGIILMNQGKTEEARR
Rat             SNMADMLYNLGLLLQENSR-FAEALHYYKLAVGSRPTLASAYLNTGIILMNQGKTEEARR
Cow             SNMADMLYNLGLLLQENSR-FAEALHYYKLAIGSRPTLASAYLNTGIILMNQGKSEEARR
Macaque         SNMADMLYNLGLLLQENSR-FSEALHYYKLAIGSRPTLASAYLNTGIILMNQGRTEEARR
Xenopus         SNMADMLYNLGLLLQENSK-FSEALHYYKLAIGSRPTLASGYLNTGIILMNQGRTEEARR
Dog             SNMADMLYNLGLLLQENSR-FAEALHYYKLAIGSRPTLASAYLNTGIILMNQGKTEEARR
Chicken         SNMADMLYNLGLLLQENNR-FSEALHYYKLAIGSRPTLASAYLNTGIILMNQGKTEEARR
CElegans        GNLADAWYNLGILTSKTSNNSNGSIHCYQMALQSRSTYAAAHLNLALLLHDAGHHMAAFS
                .*:**  ****:* .:...    ::* *::*: **.* *:.:** .::* : *:   *

Human           TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGHYEEALS---------------
Mouse           TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGRYEEALS---------------
Rat             TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGRYEEALS---------------
Cow             TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGHYEVRPEPLCPVPPPCSDLVSM
Macaque         TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGHYEEALS---------------
Xenopus         TFLKCSEIPDENLKDPNAHKSSVTSCLYNLGKLYHEQGQYEDALI---------------
Dog             TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLYHEQGHYEDALS---------------
Chicken         TFLKCSEIPDENLKDPHAHKSSVTSCLYNLGKLFHEQGHYEDALT---------------
CElegans        HLDKCLNNTGEFLKFYHVNRKTQATCAFNKGRLLQKSGNFHVAIEN--------------
                 : ** : ..* **  :.:::.: ::* :* *:* ::.*.:.

Human           -----------VYKEAIQKMPRQFAPQSLYNMMG--------------------------
Mouse           -----------VYREAIQKMPRHFAPQSLYNMMG--------------------------
Rat             -----------VYREAIQKMPRQFAPQSLYNMMG--------------------------
Cow             TSTCDLIDALSVYKEAIQKMPRQFAPQSLYNMMG--------------------------
Macaque         -----------VYKEAIQKMPRQFAPQSLYNMMG--------------------------
Xenopus         -----------VYKEAIQKMPRQFSPQSLYNMMG--------------------------
Dog             -----------VYKEAIQKMPRQFAPQSLYNMMG--------------------------
Chicken         -----------VYKEAIQKMPRQFAPQSLYNMMAFAEKPGLASHSLLRCGLLAPLQRQSK
CElegans        -----------FELALKLGGPHFEHTPSVLNSMG--------------------------
                           .       *:   . *: * *.

Human           ----------------------EAYMRLSKLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Mouse           ----------------------EAYMRLSKLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Rat             ----------------------EAYMRLSKLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Cow             ----------------------EAYMRLSMLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Macaque         ----------------------EAYMRLSKLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Xenopus         ----------------------EAYMRLNVVSEAEHWYTESLKSKPDHIPAHLTYGKLLTL
Dog             ----------------------EAYMRLSKLPEAEHWYMESLRSKTDHIPAHLTYGKLLAL
Chicken         RLGKSQMVTSTELSAAMLTAGEAYMRMSRLPEAERWYVESLRSKSDHIPAHLTYGKLLAL
CElegans        ----------------------TCYNELGDEQSAEKFFGKAIG--ENHVNSYLTMAHLKIR
                                      .* .:.   .**::: :::    :*: ::** .:*

Human           TGRKSEAEKLFLKAIELDPTKGNCYMHYGQFLLEEARLIEAAEMAKKAAELDSTEFDVVF
Mouse           TGRKSEAEKFFLKAIELDPTKGNCYMHYGQFLLEESRLTEAAEMAKKAAELDNTEFDVVF
Rat             TGRKSEAEKFFLKAIELDPTKGNCYMHYGEWHYQKQRLYEICRQMRKATDLF----ELIL
Cow             TGRKSEAEKFFLKAIELDPTKGNCYMHYGQFLLEEARLIEAAEMAKKAAELDSTEFDVVF
Macaque         TGRKSEAEKLFLKAIELDPTKGNCYMHYGEWLIAFFHVPHIYY-----------------
Xenopus         TGRKNEAERYFLKAIQLDPNKGNCYMHYGQFLLEEGRILEAAEMAKKAAELDSSEFDVVF
Dog             TGRKSEAEKFFLKAIELDPTKGNCYMHYGQFLLEEARLIEAAEMAKKAAELDSTEFDVVF
Chicken         TGRKSEAEKYFVKAIQLDPTKGNCYMHYGQFLLEESRLIEAAEMAKKAAELDNTEFDVVF
CElegans        QNRSFEVENLLRKAMTLAPESVTVLQNIALAEFHMQNYNRSLLFYRKALHLDPTHLDSLQ
                 .*. *.*. : **: * *372 .   : .       .  .
```

```
Human           NAAHMLRQASLNEAAEKYYDLAARLRPN--------------------------------
Mouse           NAAHMLRQASLNEAAEKYYDLAARLRPN--------------------------------
Rat             YPATLLKQASLNEAAEKYYDLAARLRPN--------------------------------
Cow             NAAHMLRQASLNEAAEKYYDLAARLRPN--------------------------------
Macaque         ------------------------------------------------------------
Xenopus         NAAHMLRQASLNEEAEKFYKLAAGLRQN--------------------------------
Dog             SAAHMLRQASLNEAAEKYYDLAARLRPN--------------------------------
Chicken         NAAHMLRQASLNEEAEKYYEMAAGLRPNNAFSTKLSDPRQFTVAILSNFLEIVALVRIHV
CElegans        GIANLLQQTQNHVESETFYRKVMEAQPN---------------------------------


Human           ------------------------------------YPAALMNLGAILHLNGRLQKAEANY
Mouse           ------------------------------------YPAALMNLGAILHLNGRLQKAEANY
Rat             ------------------------------------YPAALMNLGAILHLNGRLQKAEANY
Cow             ------------------------------------YPAALMNLGAILHLNGRLQKAEANY
Macaque         ------------------------------------------------------------
Xenopus         ------------------------------------YPAALMNLGAILHLNGKLEEAEYNY
Dog             ------------------------------------YPAALMNLGAILHLNGRLQKAEANY
Chicken         LFGTKYQMQHLQTPLLKRFSSATAQRTRIKEAGGKYPAALMNLGAILHLNGKLKEAEENY
CElegans        ------------------------------------SYAAHANYGAILHLNQKYDLALKEY


Human           LRALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
Mouse           LRALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
Rat             LRALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
Cow             LRALQLKPDDIITQSNLRKLWNIMEKQGLKTSKT
Macaque         ----------------------------------
Xenopus         LRALQLKPDDAITQSNLRKLWNIMEKQGLKNSKT
Dog             LRALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
Chicken         LLALQLKPDDVITQSNLRKLWNIMEKQGLKTSKT
CElegans        EIALILDPTSDVARENRKKVIRILRRKRNL----
```

ALIGNMENT OF THE DUF 1736 DOMAIN OF TMTC2 HOMOLOGS

```
Human     SISLLIFWGSSLLGARLYWMGNKPPSFSNSDNPAADSDSLLTRTLTFFYLPTKNLWLLLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYTGLLLL
Mouse     SISLLTFWGTCLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFLYLPTKNLWLLLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYSGLLLL
Rat       SISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWLLLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYSGLLLL
Cow       SISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWLLLCPDTLSFDWSMDAVPLLKTISDWRNLHTVAFYTGLLLL
Macaque   SISLLTFWGSSLLGARLYWMGNKPPSFSNSDNPAADSDSLLTRTLTFFYLPTKNLWLLLCPDTLSFDWSMDAVPLLKTVCDWRNLHTVAFYTGLLLL
Xenopus   SVGLLFAWGVILLGARFYWMGNTPPSFSNSDNPAADCEVLLTRTLTFFYLPTKNLWLLFCPDTLSFDWSMDAVPLIKTITDWRNIHTVAFYILLILL
Dog       SISLLTFWGTSLLGARLYWMGNKPPSFSNSDNPAADSDSLLARTLTFFYLPTKNLWLLLCPDTLSFDWSMDAVPLLRTVCDWRNLHTVAFYTGLLLL
Chicken   SIGLLTSWGTALLGVRLYWMGNKPPSFSNSDNPAADSDSFLTRMLTFLYLPTKNLWLLFCPDTLSFDWSMDAVPLLKAVSDWRNLHTVAFYAGLFLL
CElegans  DVTAYLISLLALCYLRLSINNFQSPKFSKNDNPIAHEPNSLTRALTFMYLPVFHLNLIVFPKTLSYDWSMDAIPKVESLIDSRIILTFIVIG-----
          .:        *   *:   .  .*.**:.*** *.    *:* ***:***. :* *:. *.***:******:* :.:: * * : *. .
```

ALIGNMENT OF THE DUF 1736 DOMAIN ACROSS MULTIPLE HOMOLOGS



References

1. Lis, H.; Sharon, N., Protein glycosylation. *Eur. J. Biochem.* **1993,** 218, (1), 1-27.
2. Schaffer. C.; Graininger, M.; Messner, P., Prokaryotic Glycosylation. *Proteomics* **2001,** 1, 248-261.
3. Law, S. K. A.; Reid, K. B. M., *Complement*. 2nd ed.; IRL Press: Oxford, 1995.
4. Ullman, C. G.; Perkins, S. J., A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *B J Letters* **1997,** 326, 929-942.
5. Lechner, J.; Wieland, F., Structure and biosynthesis of prokaryotic glycoproteins. *Ann. Rev. Biochem.* **1989,** 58, 173-194.
6. Messner, P., Prokaryotic Glycoproteins: Unexplored but Important. *J. Bacteriol.* **2004,** 186, (9), 2517-2519.
7. Schmidt, M. A.; Riley, L. W.; Benz, I., Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **2003,** 11, (12), 554-561.
8. Mescher, M. F.; Strominger, J. L.; Watson, S. W., Protein and Carbohydrate Composition of the Cell Envelope of *Halobacterium salinarium*. *J. Bacteriol.* **1974,** 120, 945–954.
9. Mescher, M. F.; Strominger, J. L., Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J. Biol. Chem.* **1976,** 251, 2005–2014.
10. Sleytr, U. B.; Messner, P.; Pum, D.; Sára, M., Crystalline bacterial cell surface layers (S-layers): from supramolecular cell structure to biomimetics and nanotechnology. *Angew. Chem. Int. Ed. Engl.* **1999,** 38, 1034–1054.
11. Sleytr, U. B., Basic and applied S-layer research: an overview. *FEMS Microbiol. Rev.* **1997,** 20, 5–12.
12. Schäffer, C.; Wugeditsch, T.; Neuninger, C.; Messner, P., Are S-Layer Glycoproteins and Lipopolysaccharides Related? *Microb. Drug Resist.* **1996,** 2, (1), 17–23.
13. Messner, P.; Christian, R.; Neuninger, C.; Schulz, G., Similarity of "core" structures in two different glycans of tyrosine- linked eubacterial S-layer glycoproteins. *J. Bacteriol.* **1995,** 177, 2188–2193.
14. Wugeditsch, T.; Zachara, N. E.; Puchberger, M.; Kosma, P.; Gooley, A. A.; Messner, P., Structural heterogeneity in the core oligosaccharide of the S-layer glycoprotein from *Aneurinibacillus thermoaerophilus* DSM 10155. *Glycobiology* **1999,** 9, 787-795.
15. Schäffer, C.; Müller N.; Christian, R.; Graninger, M.; Wugeditsch, T.; Scheberl, A.; Messner, P., Complete glycan structure of the S-layer glycoprotein of *Aneurinibacillus thermoaerophilus* GS4-97. *Glycobiology* **1999,** 9, 407–414.

16. Kärcher, U.; Schröder, H.; Haslinger, E.; Allmaier, G.; Schreiner, R.; Wieland, F.; Haselbeck, A.; König, H., Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermus fervidus*. *J. Biol. Chem.* **1993,** 268, 26821–26826.
17. Gongadze, G. M.; Kostyukova, A. S.; Miroshnichenko, M. L.; Bonch-Osmolovskaya, E. A., Regular proteinaceous layers of *Thermococcus stetteri* cell envelope. *Curr. Microbiol.* **1993,** 27, (1), 5-9.
18. Sleytr, U. B.; Sára, M.; Küpcü, Z.; Messner, P., Structural and chemical characterization of S-layers of selected strains of *Bacillus stearothermophilus* and *Desulfotomaculum nigrificans*. *Arch. Microbiol.* **1986,** 146, (1), 19-24.
19. Christian, R.; Schulz, G.; Unger, F. M.; Messner, P.; Küpcü, Z.; Sleytr, U. B., Structure of a rhamnan from the surface-layer glyco-protein of Bacillus stearothermophilus strain NRS 2004/3a. *Carbohydr. Res.* **1986,** 150, (1), 265–272.
20. Kosma, P.; Neuninger, C.; Christian, R.; Schulz, G.; Messner, P., Glycan structure of the S-layer glycoprotein of *Bacillus sp.* L420-91. *Glycoconjugate J.* **1995,** 12, (1), 99–107.
21. Kosma, P.; Wugeditsch, T.; Christian, R.; Zayni, S.; Messner, P., Glycan structure of a heptose-containing S-layer glycoprotein of *Bacillus thermoaerophilus*. *Glycobiology* **1995,** 5, 791-796.
22. Schäffer, C.; Graininger, M.; Messner, P., Prokaryotic Glycosylation. *Proteomics* **2001,** 1, 248-261.
23. Abu-Qarn, M.; Eichler, J.; Sharon, N., Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **2008,** 18, 544-550.
24. Szymanski, C. M.; Wren, B. W., Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **2005,** 3, 225-237.
25. Yurist-Doutsch, S.; Abu-Qarn, M.; Battaglia, F.; Morris, H. R.; Hitchen, P. G.; Dell, A.; Eichler, J., aglF, aglG and aglI, novel members of a gene cluster involved in the N-glycosylation of the Haloferax volcanii S-layer glycoprotein. *Mol. Microbiol.* **2008,** 69, 1234-1245.
26. Yurist-Doutsch, S.; Chaban, B.; Van Dyke, D. J.; Jarrell, K. F.; Eichler, J., Sweet to the extreme: protein glycosylation in Archaea. *Mol. Microbiol.* **2008,** 68, (5), 1079-1084.
27. Chaban, B.; Voisin, S.; Kelly, J.; Logan, S. M.; Jarrell, K. F., Identification of genes involved in the biosynthesis and attachment of Methanococcus voltae N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol. Microbiol.* **2006,** 61, 259-268.
28. Shams-Eldin, H.; Chaban, B.; Niehus, S.; Schwarz, R. T.; Jarrell, K. F., Identification of the archaeal alg7 gene homolog (N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in yeast. *J. Bacteriol.* **2008,** 190, 2217-2220.
29. Abu-Qarn, M.; Yurist-Doutsch, S.; Giordano, A.; Trauner, A.; Morris, H. R.; Hitchen, P.; Medalia, O.; Dell, A.; Eichler, J., Haloferax volcanii AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J. Mol. Biol.* **2007,** 374, 1224-1236.

30. Abu-Qarn, M.; Giordano, A.; Battaglia, F.; Trauner, A.; Morris, H. R.; Hitchen, P.; Dell, A.; Eichler, J., Identification of AglE, a second glycosyltransferase involved in N-glycosylation of the Haloferax volcanii S-layer glycoprotein. *J. Bacteriol.* **2008,** 190, (3140-3146).
31. Zeitler, R.; Hochmuth, E.; Deutzmann, R.; Sumper, M., Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from Halobacterium halobium. *Glycobiology* **1998,** 8, 1157–1164.
32. Igura, M.; Maita, N.; Kamishikiryo, J.; Yamada, M.; Obita, T.; Maenaka, K.; Kohda, D., Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **2008,** 27, 234-243.
33. Power, P. M.; Roddam, L. F.; Rutter, K.; Fitzpatrick, S. Z.; Srikhanta, Y. N.; Jennings, M. P., Genetic characterization of pilin glycosylation and phase variation in Neisseria meningitidis. *Mol. Microbiol.* **2003,** 49, 833–847.
34. Kawamura, T.; Shockman, G. D., Purification and some properties of the endogenous, autolytic N-acetylmuramoylhydrolase of Streptococcus faecium, a bacterial glycoenzyme. *J. Biol. Chem.* **1983,** 258, 9514–9521.
35. Erickson, P. R.; Herzberg, M. C., Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from Streptococcus sanguis. *J. Biol. Chem.* **1993,** 268, 23780–23783.
36. Dobos, K. M.; Khoo, K.-H.; Swiderek, K.; Brennan, P. J.; Belisle, J. T., Definition of the full extent of glycosylation of the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *J. Bacteriol.* **1996,** 178, 2498–2506.
37. Benz, I.; Schmidt, M. A., Glycosylation with heptose residues mediated by the aah gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol.* **2001,** 40, 1403–1413.
38. Moormann, C.; Benz, I.; Schmidt, M. A., Functional substitution of the autotransporter adhesin heptosyltransferase (AAH) of the AIDA system by the TibC protein of enterotoxigenic Escherichia coli strains. *Infect. Immun.* **2002,** 70, (2264–2270).
39. Szymanski, C. M.; Logan, S. M.; Linton, D.; Wren, B. W., Campylobacter-– a tale of two protein glycosylation systems. *Trends Microbiol.* **2003,** 11, 233-238.
40. Kowarik, M. et. al, Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **2006,** 25, 1957–1966.
41. Linton, D.; Dorrell, N.; Hitchen, P. G.; Amber, S.; Karlyshev, A. V.; Morris, H. R.; Dell, A.; Valvano, M. A.; Aebi, M.; Wren, B. W., Functional analysis of the *Campylobacter jejuni N*-linked protein glycosylation pathway. *Mol. Microbiol.* **2005,** 55, 1695–1703.
42. Virji, M., Glycans in Meningococcal pathogenesis and the enigma of the molecular decorations of Neisserial pili. In *Glycomicrobiology*, Doyle, R. J., Ed. Kluwer Academic/Plenum Publishers: 2000; p 31–65.
43. Linton, D., Allan, E., Karlyshev, A.V., Cronshaw, A.D., Wren, B.W., Identification of N-acetylgalactosaminecontaining glycoproteins PEB3 and CgpA in Campylobacter jejuni. *Mol. Microbiol.* **2002,** 43, 497–508.

44. Wacker, M.; Linton, D.; Hitchen, P. G.; Nita-Lazar, M.; Haslam, S. M.; North, S. J.; Panico, M.; Morris, H. R.; Dell, A.; Wren, B. W.; Aebi, M., *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **2002,** 298, 1790–1793.
45. Szymanski, C. M. et. al, Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol. Microbiol.* **1999,** 32, 1022–1030.
46. Burda, P.; Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* **1999,** 1426, 239-257.
47. Yan, Q.; Lennarz, W. J., Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* **2002,** 277, 47692–47700.
48. Kowarik, M.; Numao, S.; Feldman, M. F.; Schulz, B. L.; Callewaert, N.; Kiermaier, E.; Catrein, I.; Aebi, M., N-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase. *Science* **2006,** 314, 1148-1149.
49. Thibault, P.; Logan, S. M.; Kelly, J. F.; Brisson, J. R.; Ewing, C. P.; Trust, T. J.; Guerry, P., Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *J. Biol. Chem.* **2001,** 276, 34862–34870.
50. Samuel, G.; Reeves, P., Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr. Res.* **2003,** 338, 2503-2519.
51. Johnson, D. A.; Liu, H.-w., Mechanisms and pathways from recent deoxysugar biosynthesis research. *Curr. Opin. Chem. Biol.* **1998,** 2, (5), 642-649.
52. Trezfer, A.; Salas, J. A.; Bechthold, A., Genes and enzymes involved in deoxysugar biosynthesis in bacteria. *Nat. Prod. Rep.* **1999,** 16, 283-299.
53. Yamamoto, S.; Miyake, K.; Koike, Y.; Watanabe, M.; Machida, Y.; Ohta, M.; Iijima, S. J., Molecular Characterization of Type-Specific Capsular Polysaccharide Biosynthesis Genes of *Streptococcus agalactiae* Type Ia. *J. Bacteriol.* **1999,** 181, (17), 5176-5184.
54. Reeves, P. R., Biosynthesis and assembly of lipopolysaccharide. In *Bacterial cell wall*, Neuberger, A.; van Deenen, L. L. M., Eds. Elsevier: Amsterdam, 1994; p 281–314.
55. Upreti, R. K.; Kumar, M.; Shankar, V., Bacterial glycoproteins: functions, biosynthesis and applications. *Proteomics* **2003,** 3, 363–379.
56. Banerjee, A.; Gosh, S. K., The role of pilin glycan in neisserial pathogenesis. *Mol. Cell. Biochem.* **2003,** 253, 179–190.
57. Castric, P.; Cassels, F. J.; Carlson, R. W., Structural characterization of the Pseudomonas aeruginosa 1244 pilin glycan. *J. Biol. Chem.* **2001,** 276, 26479–26485.
58. Stimson, E.; Virji, M.; Makepeace, K.; Dell, A.; Morris, H. R.; Payne, G.; Saunders, J. R.; Jennings, M. P.; Barker, S.; Panico, M.; Blench, I.; Moxon, E. R., Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **1995,** 17, 1201–1214.
59. Jarrell, K. F.; Bayley, D. P.; S., K. A., The archaeal flagellum: a unique motility structure. *J. Bacteriol.* **1996,** 178, 5057–5064.

60. Sumper, M.; Wieland, F. T., *Bacterial glycoproteins*. Elsevier: Amsterdam, 1995; Vol. 29a, p 455–473.
61. Schirm, M.; Soo, E. C.; Aubry, A. J.; Austin, J.; Thibault, P.; Logan, S. M., Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. *Mol. Microbiol.* **2003,** 48, 1579–1592.
62. Helenius, A.; Aebi, M., Intracellular Functions of N-Linked Glycans. *Science* **2001,** 291, 2364-2369.
63. Feng, L.; Tao, J.; Hongjie, G.; Xu, J.; Li, Y.; Rezwan, F.; Reeves, P.; Wang, L., Structure of the Shigella dysenteriae 7 O antigen gene cluster and identification of its antigen specific genes. *Microb. Path.* **2004,** 36, 109-115.
64. D'Souza, J.; Wang, L.; Reeves, P., Sequence of the Escherichia coli O26 O antigen gene cluster and identification of O26 specific genes. *Gene* **2002,** 297, 123-127.
65. Lamb, S., Wright, G., Accessorizing natural products: Adding to nature's toolbox. *Proc. Nat. Acad. Sci.* **2005,** 102, 519-520.
66. Khosla, C.; Keasling, J. D., Metabolic engineering for drug discovery and development. *Nat. Rev. Drug Discovery* **2003,** 2, (12), 1019-1025.
67. Walsh, C. T., Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. *Science* **2004,** 303, (5665), 1805-1810.
68. Zhang, H.; White-Phillip, J. A.; Melancon, C. E.; Kwon, H.-j.; Yu, W.-l.; Liu, H.-w., Elucidation of the Kijanimicin Gene Cluster: Insights into the Biosynthesis of Spirotetronate Antibiotics and Nitrosugars. *J. Am. Chem, Soc.* **2007,** 127, 14670-14683.
69. Luzhetskyy, A.; Vente, A.; Bechthold, A., Glycosyltransferases involved in the biosynthesis of biologically active natural products that contain oligosaccharides. *Mol. BioSyst.* **2005,** 1, 117-126.
70. Weymouth-Wilson, A. C., The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep.* **1997,** 14, 99-110.
71. Luzhetskyy, A.; Vente, A.; Bechthold, A., Glycosyltransferases involved in the biosynthesis of biologically active natural products that contain oligosaccharides. *Mol. BioSyst.* **2005,** 1, 117-126.
72. Walsh, C.; Freel-Meyers, C. L.; Losey, H. C., Antibiotic Glycosyltransferases: Antibiotic Maturation and Prospects for Reprogramming. *J. Med. Chem.* **2003,** 46, 3425-3436.
73. Bililign, T.; Griffith, B. R.; Thorson, J. S., Structure, activity, synthesis and biosynthesis of aryl-C-glycosides. *Nat. Prod. Rep.* **2005,** 22, 742-760.
74. Hantke, K.; Nicholson, G.; Rabsch, W.; Winkelmann, G., Bioinorganic Chemistry Special Feature: Salmochelins, siderophores of Salmonella enterica and uropathogenic Escherichia coli strains, are recognized by the outer membrane receptor IroN. *Proc. Nat. Acad. Sci. USA* **2003,** 100, (7), 3677–3682.
75. Fischbach, M. A.; Lin, H.; Liu, D. R.; Walsh, C. T., In vitro characterization of IroB, a pathogen-associated C-glycosyltransferase. *Proc. Nat. Acad. Sci. USA* **2005,** 102, (3), 571-576.
76. Hu, Y.; Walker, S., Remarkable Structural Similarities between Diverse Glycosyltransferases. *Chem Biol* **2002,** 9, 1287–1296.

77. Pedersen, L. C.; Tsuchida, K.; Kitagawa, H.; Sugahara, K.; Rini, J., *EMBO J.* **2000,** 19, 5269–5280.
78. Thibodeaux, C. J.; Melancon, C. E.; Liu, H.-w., Natural Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chem. Int. Ed. Engl.* **2008,** 47, (51), 9814-9859.
79. Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G., Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **2008,** 77, 521-555.
80. Coutinho, P. M.; Deleury, E.; Davies G. J.; Henrissat, B., An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **2003,** 328, (2), 307-317.
81. Campbell, R. E.; Mosimann, S. C.; Tanner, M. E.; Strynadka, N. C. J., The Structure of UDP-N-Acetylglucosamine 2-Epimerase Reveals Homology to Phosphoglycosyl Transferases. *Biochemistry* **2000,** 39, 14993–15001.
82. Murray, B. W.; Takayama, S.; Schultz, J.; Wong, C.-H., Mechanism and Specificity of Human α-1,3-Fucosyltransferase V. *Biochemistry* **1996,** 35, (34), 11183-11195.
83. Zechel, D. L.; Withers, S. G., Glycosidase Mechanisms: Anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **2000,** 33, 11-18.
84. Durr, C.; Hoffmeister, D.; Wohlert, S.-E.; Ichinose, K.; Weber, M.; von Mulert, U.; Thorson, J. S.; Bechthold, A., The glycosyltransferase UrdGT2 catalyzes both C- and O-glycosidic sugar transfers. *Angew. Chem. Int. Ed. Engl.* **2004,** 43, 2962-2965.
85. Mittler, M.; Bechthold, A.; Schulz, G. E., Structure and Action of the C–C Bond-forming Glycosyltransferase UrdGT2 Involved in the Biosynthesis of the Antibiotic Urdamycin. *J. Mol. Biol.* **2007,** 372, (1), 67-76.
86. Van den Steen, P. E.; Rudd, P. M.; Wormald, M. R.; Dwek, R. A.; Opdenakker, G., *O*-Linked Glycosylation in Focus. *Trends in Glycosci. and Glycotech.* **2000,** 12, (63), 35-49.
87. Varki, A., Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **1993,** 3, (2), 97-130.
88. Drickamer, K.; Taylor, M. E., Evolving views of protein glycosylation. *TIBS* **1998**, 321-324.
89. Schachter, H., The clinical relevance of glycobiology. *The Journal of Clinical Investigation* **2001,** 108, (11), 1579–1582.
90. Spiro, R. G., Protein glycosylation: nature, distribution, enzymatic formation and disease implications of glycopeptide bonds. *Glycobiology* **2002,** 12, (4), 43R-56R.
91. Lehle, L., Strahl, S., and Tanner, W., Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angew. Chem. Int. Ed. Engl.* **2006,** 45, 6802-6818.
92. Rose, M. C.; Voynow, J. A., Respiratory Tract Mucin Genes and Mucin Glycoproteins in Health and Disease. *Physiol Rev* **2006,** 86, 245-278.
93. Stagljar, I.; te Heesen, S.; Aebi, M., New phenotype of mutations deficient in glucosylation of the lipid-linked oligosaccharide: cloning of the ALG8 locus. *Proc. Nat. Acad. Sci. USA* **1994,** 91, 5977-5981.

94. Manya, H.; Takehiro Suzuki, T.; Akasaka-Manya, K.; Ishida, H.-K.; Mizuno, M.; Suzuki, Y.; Inazu, T.; Dohmae, N.; Endo, T., Regulation of Mammalian Protein O-Mannosylation: PREFERENTIAL AMINO ACID SEQUENCE FOR O-MANNOSE MODIFICATION. *J. Biol. Chem.* **2007,** 282, 20200-20206.
95. Manya, H.; Chiba, A.; Yoshida, A.; Wang, X.; Chiba, Y.; Jigami, Y.; Margolis, R. U.; Endo, T., Demonstration of mammalian protein *O*-mannosyltransferase activity: Coexpression of POMT1 and POMT2 required for enzymatic activity. *Proc. Nat. Acad. Sci. USA* **2004,** 101, (2), 500-505.
96. Willer, T.; Amselgruber, W.; Deutzmann, R.; Strahl, S., Characterization of POMT2, a novel member of the PMT protein O-mannosyltransferase family specifically localized to the acrosome of mammalian spermatids. *Glycobiology* **2002,** 12, (771–783).
97. Kornfeld, R.; Kornfeld, S., Assembly of asparagine-linked oligosaccharides. *Ann. Rev. Biochem.* **1985,** 54, 631-664.
98. Weerapana, E.; Imperiali, B., Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **2006,** 16, (6), 91-101.
99. Bugg, T. D.; Brandish, P. E., From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis. *FEMS Microbiol. Lett.* **1994,** 119, 255-262.
100. Geremia, R. A.; Petroni, E. A.; Ielpi, L.; Henrissat, B., Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic alpha-mannosyltransferases. *Biochem. J.* **1996,** 318, 133-138.
101. Takeda, J.; Kinoshita, T., GPI-anchor biosynthesis. *Trends Biochem. Sci.* **1995,** 20, 367-371.
102. Whitfield, C., Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol.* **1995,** 3, 178-185.
103. Kandler, O.; Koenig, H., *Cell envelopes of archaea: structure and chemistry*. Elsevier: Amsterdam, 1993; Vol. 26, p 223-259.
104. Spirig, U.; Glavas, M.; Bodmer, D.; Reiss, G.; Burda, P.; Lippuner, V.; Te Heesen, S.; Aebi, M., The STT3 protein is a component of the yeast oligosaccharyltransferase complex. *Mol. Gen. Genet.* **1997,** 256, 628-637.
105. Snider, M. D.; Rogers, O. C., Transmembrane movement of oligosaccharide-lipids during glycoprotein synthesis. *Cell* **1984,** 36, (3), 753-761.
106. Snider, M. D.; Robbins, P. W., Transmembrane organization of protein glycosylation. Mature oligosaccharide-lipid is located on the luminal side of microsomes from Chinese hamster ovary cells. *J. Biol. Chem.* **1982,** 257, 6796-6801.
107. Bickel, T.; Lehle, L.; Schwarz, M.; Aebi, M.; Jakob, C. A., Biosynthesis of lipid-linked oligosaccharides in *Saccharomyces cerevisiae*: Alg13p and Alg14p form a complex required for the formation of GlcNAc(2)-PP-dolichol. *J. Biol. Chem.* **2005,** 280, 34500-34506.
108. Gao, X. D.; Tachikawa, H.; Sato, T.; Jigami, Y.; Dean, N., Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation. *J. Biol. Chem.* **2005,** 280, 36254-36262.

109. O' Reilly, M. K.; Zhang, G.; Imperiali, B., In Vitro Evidence for the Dual Function of Alg2 and Alg11: Essential Mannosyltransferases in N-Linked Glycoprotein Biosynthesis. *Biochemistry* **2006,** 45, (31), 9595-9603.
110. Dickinson, J. R., Biochemical and genetic studies on the function of, and relationship between the PGI1- and CDC30- encoded phosphoglucose isomerases in *Saccharomyces cerevisiae*. *J. Gen. Microbiol* **1991,** 137, 765-770.
111. Watzele, G.; Tanner, W., Cloning of the glutamine:fructose-6-phosphate amidotransferase gene from yeast. Pheromonal regulation of its transcription. *J. Biol. Chem.* **1989,** 264, 8753-8758.
112. Hofmann, M.; Boles, E.; Zimmermann, F. K., Characterization of the essential yeast gene encoding *N*-acetylglucosaminephosphate mutase. *Eur. J. Biochem.* **1994,** 221, 741-747.
113. Smith, D. J.; Proudfoot, A.; Friedli, L.; Klig, L. S.; Paravicini, G.; Payton, M. A., PM140, an intron-containing gene required for early steps in yeast mannosylation. *Mol. Cell. Biol.* **1992,** 12, 2924-2930.
114. Kepes, F.; Schekman, R., The yeast SEC53 gene encodes phosphomannomutase. *J. Biol. Chem.* **1988,** 263, 9155-9161.
115. Hashimoto, H.; Sakakibara, A.; Yamasaki, M.; Yoda, K., Saccharomyces cerevisiae VIG9 encodes GDP-mannose pyrophosphorylase, which is essential for protein glycosylation. *J. Biol. Chem.* **1997,** 272, 16308-16314.
116. Crick, D. C.; Rush, J. S.; Waechter, C. J., Characterization and localization of a long-chain isoprenyltransferase activity in porcine brain: proposed role in the biosynthesis of dolichyl phosphate. *J. Neurochem.* **1991,** 57, 1354-1362.
117. Hubbard, S. C.; Robbins, P. W., Synthesis of *N*-linked oligosaccharides of glycoproteins. *J. Biol. Chem.* **1980,** 255, 11782-11793.
118. Crick, D. C.; Waechter, C. J., Long-chain cis-isoprenyltransferase activity is induced early in the developmental program for protein *N*-glycosylation in embryonic rat brain cells. *J. Neurochem.* **1994,** 62, 247-256.
119. Carlberg, M.; Dricu, A.; Blegen, H.; Wang, M.; Hjertman, M.; Zickert, P.; Hoog, A.; Larsson, O., Mevalonic acid is limiting for *N*-linked glycosylation and translocation of the insulinlike growth factor-1 receptor to the cell surface - Evidence for a new link between 3-hydroxy-3-methylglutaryl-coenzyme A reductase and cell growth. *J. Biol. Chem.* **1996,** 271, 17453-17462.
120. Eigen, M., *Quantum Statistical Mechanics in the Natural Sciences*. Plenum Press: New York, NY, 1974; p 37-61.
121. Oh, D.; Hopper, J. E., Transcription of a yeast phosphoglucomutase isozyme gene is galactose inducible and glucose repressible. *Mol. Cell. Biol.* **1990,** 10, 1415-1422.
122. Boles, E.; Liebetrau, W.; Hofmann, M.; Zimmermann, F. K., A family of hexosephosphate mutases in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **1994,** 220, 83-96.
123. te Heesen, S.; Lehle, L.; Weissmann, A.; Aebi, M., Isolation of the ALG5 locus encoding the UDP-glucose:dolichyl-phosphate glucosyltransferase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **1994,** 224, 71-79.

124. Schutzbach, J. S.; Zimmerman, J. W.; Forsee, W. T., The Purification and Characterization of Recombinant Yeast Dolichyl-phosphate-mannose Synthase. *J. Biol. Chem.* **1993,** 268, (32), 24190-24196.
125. Orlean, P.; Albright, C.; Robbins, P. W., Cloning and Sequencing of the Yeast Gene for Dolichol Phosphate Mannose Synthase, an Essential Protein. *J. Biol. Chem.* **1988,** 263, (33), 17499-17507.
126. Doucey, M.-A.; Hess, D.; Cacan, R.; Hofsteenge, J., Protein C-Mannosylation Is Enzyme-catalysed and Uses Dolichyl-Phosphate-Mannose as a Precursor. *Mol. Biol. Cell* **1998,** 9, 291-300.
127. Aebi, M.; Gassenhuber, J.; Domdey, H.; Heesen, S. T., Cloning and characterization of the ALG3 gene of *Saccharomyces cerevisiae*. *Glycobiology* **1996,** 6, 439-444.
128. Rearick, J. I.; Fujimoto, K.; Kornfeld, S., Identification of the mannosyl donors involved in the synthesis of lipid-linked oligosaccharides. *J. Biol. Chem.* **1981,** 256, 3762-3769.
129. Burda, P.; te Heesen, S.; Brachat, A.; Wach, A.; Dusterhoft, A.; Aebi, M., Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of *Saccharomyces cerevisiae*: Identification of the *ALG9* gene encoding a putative mannosyl transferase. *Proc. Nat. Acad. Sci. USA* **1996,** 93, 7160-7165.
130. Reiss, G.; te Heesen, S.; Zimmerman, J.; Robbins, P. W.; Aebi, M., Isolation of the ALG6 locus of *Saccharomyces cerevisiae* required for glucosylation in the *N*-linked glycosylation pathway. *Glycobiology* **1996,** 6, 493-498.
131. Helenius, A.; Trombetta, E. S.; Hebert, D. N.; Simons, J. F., Calnexin, calreticulin and the folding of glycoproteins. *Trends Cell Biol.* **1997,** 7, 193-200.
132. Zimmerman, J. W.; Specht, C. A.; Cazares, B. X.; Robbins, P. W., The isolation of a Dol-P-Man synthase from *Ustilago maydis* that functions in *Saccharomyces cerevisiae*. *Yeast* **1996,** 12, 765-771.
133. Li, G.; Yan, Q.; Nita-Lazar, A.; Haltiwanger, R. S.; Lennarz, W. J., Studies on the *N*-Glycosylation of the Subunits of Oligosaccharyl Transferase in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **2005,** 280, (3), 1864-1871.
134. Petrescu, A. J.; Milac, A. L.; Petrescu, S. M.; Dwek, R. A.; Wormald, M. R., Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **2004,** 14, 103-114.
135. Yan, A.; Ahmed, E.; Yan, Q.; Lennarz, W. J., New findings on interactions among the yeast oligosaccharyl transferase subunits using a chemical cross-linker. *J. Biol. Chem.* **2003,** 278, 33078–33087.
136. Dempski Jr. R. E.; Imperiali, B., Heterologous expression and biophysical characterization of soluble oligosaccharyl transferase subunits. *Arch Biochem Biophys* **2004,** 431, 63–70.
137. Lennarz, W. J., Studies on oligosaccharyl transferase in yeast. *Acta Biochemica Polonica* **2007,** 54, (4), 1-5.

138. Wilson, C. M.; Kraft, C.; Duggan, C.; Ismail, N.; Crawshaw, S. G.; High, S., Ribophorin I associates with a subset of membrane proteins after their integration at the Sec61 translocon. *J. Biol. Chem.* **2005,** 280, 4195–4206.
139. Yan, Q.; Lennarz, W. J., Studies on the function of oligosaccharyl transferase subunits: A glycosylatable photoprobe binds to the luminal domain of Ost1p. *Proc. Nat. Acad. Sci. USA* **2002,** 99, 15994–15999.
140. Reiss, G.; te Heesen, S.; Gilmore, R.; Zufferey, R.; Aebi, M., A specific screen for oligosaccharyltransferase mutations identifies the 9 kDa OST5 protein required for optimal activity *in vivo* and *in vitro*. *EMBO J.* **1997,** 16, 1164–1172.
141. Yan, A.; Wu, E.; Lennarz, W. J., Studies of yeast oligosaccharyl transferase subunits using the split-ubiquitin system: Topological features and *in vivo* interactions. *Proc. Nat. Acad. Sci. USA* **2005,** 102, 7121–7126.
142. te Heesen, S.; Knauer, R.; Lehle, L.; Aebi, M., Yeast Wbp1p and Swp1p form a protein complex essential for oligosaccharyl transferase activity. *EMBO J.* **1993,** 12, 279–284.
143. Chavan, M.; Yan, A.; Lennarz, W. J., Subunits of the translocon interact with components of the oligosaccharyl transferase complex. *J. Biol. Chem.* **2005,** 280, 22917–22924.
144. Li, G.; Yan, Q.; Oen H.O.; Lennarz, W. J., A specific segment of the transmembrane domain of Wbp1p is essential for its incorporation into the oligosaccharyl transferase complex. *Biochemistry* **2003,** 42, 11032–11039.
145. Bause, E.; Wesemann, M.; Bartoschek, A.; Breuer, W., Epoxyethylglycyl peptides as inhibitors of oligosaccharyltransferase:Double-labelling of the active site. *Biochem. J.* **1997,** 322, 95–102.
146. Pathak, R.; Hendrickson, T. L.; Imperiali, B., Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry* **1995,** 34, 4179–4185.
147. Chi, J. H.; Roos, J.; Dean, N., The OST4 gene of Saccharomyces cerevisiae encodes an unusually small protein required for normal levels of oligosaccharyltransferase activity. *J. Biol. Chem.* **1996,** 271, 3132-3140.
148. Kim, H.; Park, H.; Montalvo, L.; Lennarz, W. J., Studies on the role of the hydrophobic domain of Ost4p in interactions with other subunits of yeast oligosaccharyltransferase. *Proc. Nat. Acad. Sci. USA* **2000,** 97, 1516-1520.
149. Kim, H.; Yan, Q.; Von Heijne, G.; Caputo, G. A.; Lennarz, W. J., Determination of the membrane topology of Ost4p and its subunit interactions in the oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci. USA* **2003,** 100, 7460-7464.
150. Spirig, U.; Bodmer, D.; Wacker, M.; Burda, P.; M, A., The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **2005,** 15, 1396-1406.
151. Yan, A.; Lennarz, W. J., Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* **2005,** 15, 1407-1415.

152. Knauer, R.; Lehle, L., The oligosaccharyltransferase complex from Saccharomyces cerevisiae. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. *J. Biol. Chem.* **1999,** 274, 17249–17256.
153. Yan, Q.; Prestwich, G. D.; Lennarz, W. J., The Ost1p subunit of yeast oligosaccharyl transferase recognizes the peptide glycosylation site sequence, -Asn-X-Ser/Thr-. *J. Biol. Chem.* **1999,** 274, 5021-5025.
154. Glover, K. J.; Weerapana, E.; Numao, S.; Imperiali, B., Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol.* **2005,** 12, 1311-1315.
155. Nilsson, I.; Kelleher, D. J.; Miao, Y.; Shao, Y.; Kreibich, G.; Gilmore, R.; von Heijne, G.; Johnson, A. E., Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J. Cell. Biol.* **2003,** 161, 715–725.
156. Imperiali, B.; O'Connor, S. E., Effect of *N*-linked glycosylation on glycopeptide and glycoprotein structure. *Curr. Opin. Chem. Biol.* **1999,** 3, (6), 643-649.
157. Wormald, M. R.; Dwek, R. A., Glycoproteins: glycan presentation and protein-fold stability. *Structure Fold. Des.* **1999,** 7, (7), R155-160.
158. Beintema, J. J., Do asparagine-linked carbohydrate chains in glycoproteins have a preference for β-bends? *Biosci. Rep.* **1986,** 6, 709-714.
159. O'Connor, S. E.; Imperiali, B., Modulation of protein structure and function by asparagine-linked glycosylation. *Chem. Biol.* **1996,** 3, (10), 803-812.
160. Kern, G.; Kern, D.; Jaenicke, R.; Seckler, R., Kinetics of folding and association of differently glycosylated variants of invertase from *Saccharomyces cerevisiae*. *Protein Sci.* **1993,** 2, (11), 1862-1868.
161. Helenius, A., How *N*-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* **1994,** 5, 253-265.
162. Olden, K.; Parent, J. B.; White, S. L., Carbohydrate moieties of glycoproteins- a re-evaluation of their function. *Biochim. Biophys. Acta* **1982,** 650, (4), 209-232.
163. Jaenicke, R., Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry* **1991,** 30, (13), 3147-3161.
164. Zapun, A.; Jakob, C. A.; Thomas, D. Y.; Bergeron, J. J., Protein folding in a specialized compartment: the endoplasmic reticulum. *Structure Fold. Des.* **1999,** 7, 173-182.
165. Parodi, A. J., Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.* **2000,** 348, 1-13.
166. Ou, W.-J.; Cameron, P. H.; Thomas, D. Y.; Bergeron, J. J. M., Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature* **1993,** 364, 771-776.
167. Hammond, C.; Braakman, I.; Helenius, A., Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Nat. Acad. Sci. USA* **1994,** 91, (3), 913-917.
168. Bouvier, M.; Stafford, W. F., Probing the three-dimensional structure of human calreticulin. *Biochemistry* **2000,** 39, (48), 14950-14959.

169. Ellgaard, L.; et al., Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment, *FEBS Lett.* **2000,** 488, 69.
170. Ellgaard, L.; Riek, R.; Herrmann, T.; Güntert, P.; Braun, D.; Helenius, A.; Wüthrich, K., NMR structure of the calreticulin P-domain. *Proc. Nat. Acad. Sci. USA* **2001,** 98, (6), 3133-3138.
171. Oliver, J. D.; van der Wal, F. J.; Bulleid, N. J.; High, S., Interaction of the Thiol-Dependent Reductase ERp57 with Nascent Glycoproteins. *Science* **1997,** 275, 86-88.
172. Molinari, M.; Helenius, A., Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature* **1999,** 402, 90-93.
173. Trombetta, S.; Bosch, M.; Parodi, A. J., Glucosylation of glycoproteins by mammalian, plant fungal and trypanosomatid protozoa microsomal membranes. *Biochemistry* **1989,** 28, 8108-8116.
174. Fernandez, F. S.; Trombetta, S. E.; Hellman, U.; Parodi, A. J., Purification to homogeneity of UDP-glucose:glycoprotein glucosyltransferase from *Schizosaccharimyces cerevisae*. *J. Biol. Chem.* **1994,** 269, 30701–30706.
175. Perker, C. G.; Fessler, L. I.; Nelson, R. E.; Fessler, J. H., Drosophila UDP-glucose:glycoprotein glucosyltransferase: sequence and characterization of an enzyme that distinguishes between denatured and native proteins. *EMBO J.* **1995,** 14, 1294–1303.
176. Arnold, S. M.; Kaufman, R. J., The noncatalytic portion of human UDP-glucose:glycoprotein glucosyltransferase I confers UDP-glucose binding and transferase function to the catalytic domain. *J. Biol. Chem.* **2003,** 278, 43320–43328.
177. Arnold, S. M.; Fessler, L. I.; Fessler, J. H.; Kaufman, R. J., Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity. *Biochemistry* **2000,** 39, 2149-2163.
178. Totani, K.; Ihara, Y.; Tsuiimoto, T.; Matsuo, I.; Ito, Y., The Recognition Motif of the Glycoprotein-Folding Sensor Enzyme UDP-Glc:Glycoprotein Glucosyltransferase. *Biochemistry* **2009,** 48, (13), 2933-2940.
179. Sousa, M. C.; Ferrero-Garcia, M. A.; Parodi, A. J., Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry* **1992,** 31, (1), 97-105.
180. Cannon, K. S.; Helenius, A., Trimming and Readdition of Glucose to *N*-Linked Oligosaccharides Determines Calnexin Association of a Substrate Glycoprotein in Living Cells. *J. Biol. Chem.* **1999,** 274, 7537-7544.
181. Trombetta, E. S.; Helenius, A., Glycoprotein reglucosylation and nucleotide sugar utilization in the secretory pathway: identification of a nucleoside diphosphatase in the endoplasmic reticulum. *EMBO J.* **1999,** 18, (12), 3282-3292.
182. Ritter, C.; Helenius, A., Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *Nat. Struct. Biol.* **2000,** 7, 278-280.
183. Caramelo, J. J.; Castro, O. A.; Alonso, L. G.; Prat-Gay, G.; Parodi, A. J., UDP-Glc:glycoprotein glucosyltransferase recognizes structured and solvent accessible hydrophobic patches on molten globule-like folding intermediates. *Proc. Nat. Acad. Sci. USA* **2003,** 100, 86-91.

184. Caramelo, J. J.; Castro, O. A.; Prat-Gay, G.; Parodi, A. J., The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. *J. Biol. Chem.* **2004,** 279, 46280–46285.
185. Taylor, S. C.; Ferguson, A. D.; Bergeron, J. J. M.; Thomas, D. Y., The ER protein folding sensor UDP-glucose glycoprotein-glucosyltransferase modifies substrates distant to local changes in glycoprotein conformation. *Nat. Struct. Mol. Bio.* **2004,** 11, 128–134.
186. Ritter, C.; Quirin, K.; Kowarik, M.; Helenius, A., Minor folding defects trigger local modification of glycoproteins by ER folding sensor GT. *EMBO J.* **2005,** 24, 1730–1738.
187. Taylor, S. C.; Thibault, P.; Tessier, D. C.; Bergeron, J. J. M.; Thomas, D. Y., Glycopeptide specificity of the secretory protein folding sensor UDP-glucose glycoprotein:glucosyltransferase. *EMBO Rep.* **2003,** 4, 405-411.
188. Mesaeli, N.; Nakamura, K.; Zvaritch, E.; Dickie, P.; Dziak, E.; Krause, K.-H.; Opas, M.; MacLennan, D. H.; Michalak, M., Calreticulin Is Essential for Cardiac Development. *J. Cell. Biol.* **1999,** 144, (5), 857-868.
189. De Praeter, C. M.; Gerwig, G. J.; Bause, E.; Nuytinck, L. K.; Vliegenthart, J. F. G.; Breuer, W.; Kamerling, J. P.; Espeel, M. F.; R. Martin, J.-J. R.; De Paepe. A. M.; al., e., A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency. *Am. J. Hum. Genet.* **2000,** 66, (6), 1744-1756.
190. Molinari, M., N-Glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **2007,** 3, 313-320.
191. Plemper, R. K.; Wolf, D. H., Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem. Sci.* **1999,** 24, (7), 266-270.
192. Bonifacino, J. S.; Weissman, A. M., Ubiquitin and the control of protein fate in the secretory and endocytic pathways. *Annu. Rev. Cell Dev. Biol.* **1998,** 14, 19-57.
193. Kopito, R. R., ER quality control: The cytoplasmic connection. *Cell* **1997,** 88, (4), 427-430.
194. Liu, Y.; Choudhury, P.; Cabral, C. M.; Sifers, R. N., Intracellular Disposal of Incompletely Folded Human α1-Antitrypsin Involves Release from Calnexin and Post-translational Trimming of Asparagine-linked Oligosaccharides. *J. Biol. Chem.* **1997,** 272, 7946-7951.
195. Jakob, C. A.; Burda, P.; Roth, J.; Aebi, M., Degradation of Misfolded Endoplasmic Reticulum Glycoproteins in *Saccharomyces cerevisiae* Is Determined by a Specific Oligosaccharide Structure. *J. Cell. Biol.* **1998,** 142, (5), 1223-1223.
196. Brockhausen, I.; Schachter, H.; Stanley, P., *O*-GalNac Glycans (9). In *Essentials of Glycobiology*, 2 ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C.; Hart, G. W.; Etzler, M. E., Eds. Coldspring Harbor Laboratory Press: Coldspring Harbor, NY, 2009; p 784.
197. Elhammer, A. P.; Poorman, R. A.; Brown, E.; Maggiora, L. L.; Hoogerheide, J. G.; Kézdy, F. J., The specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides. *J. Biol. Chem.* **1993,** 268, 10029–10038.

198. Clausen, H.; Bennett, E. P., A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* **1996,** 6, 635–646.
199. Otvos, L. J.; Krivulka, G. R.; Urge, L.; Szendrei, G. I.; Nagy, L.; Xiang, Z. Q.; Ertl, H. C., Comparison of the effects of amino acid substitutions and β-N- vs. α-O-glycosylation on the T-cell stimulatory activity and conformation of an epitope on the rabies virus glycoprotein. *Biochim. Biophys. Acta* **1995,** 1267, (1), 55–64.
200. Hart, G. W.; Kreppel, L. K.; Comer, F. I.; Arnold, C. S.; Snow, D. M.; Ye, Z. Y.; Cheng, X. G.; DellaManna, D.; Caine, D. S.; Earles, B. J.; Akimoto, Y.; Cole, R. N.; K., H. B., O-GlcNAcylation of key nuclear and cytoskeletal proteins: Reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology* **1996,** 6, 711-716.
201. Hart, G. W., Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.* **1997,** 66, 315-335.
202. Torres, C.; Hart, G. W., Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. *J. Biol. Chem.* **1984,** 259, 3308-3317.
203. Hart, G. W.; Haltiwanger, R. S.; Holt, G. D.; Kelly, W. G., Glycosylation in the nucleus and cytoplasm. *Annu. Rev. Biochem.* **1989,** 58, 841-874.
204. Hart, G. W.; Akimoto, Y., The O-GlcNac Modification (18). In *Essentials of Glycobiology*, 2 ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C.; Hart, G. W.; Etzler, M. E., Eds. Coldspring Harbor Press: Coldspring Harbor, NY, 2009; p 784.
205. O'Donnell, N.; Zachara, N. E.; Hart, G. W.; Marth, J., Ogt-Dependent X-Chromosome-Linked Protein Glycosylation Is a Requisite Modification in Somatic Cell Function and Embryo Viability. *Mol. Cell. Bio.* **2004,** 24, (4), 1680-1690.
206. Kelly, W. G.; Dahmus, M. E.; Hart, G. W., RNA Polymerase II is a Glycoprotein Modification of the Cooh-terminal Domain by O-glcnac. *J. Biol. Chem.* **1993,** 268, 10416-10424.
207. Jinek, M.; Rehwinkel, J.; Lazarus, B. D.; Izaurralde, E.; Hanover, J. A.; Conti, E., The superhelical TPR-repeat domain of O-linkef GlcNAc transferase exhibits structural similarities to importin α. *Nat. Struct. Mol. Bio.* **2004,** 11, (10), 1001-1007.
208. Blatch, G. L.; Lassle, M., The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays* **1999,** 21, 932-939.
209. Kreppel, L. K.; Hart, G. W., Regulation of a Cytosolic and Nuclear *O*-GlcNAc Transferase. *J. Biol. Chem.* **1999,** 274, (45), 32015-32022.
210. Freeze, H. H.; Haltiwanger, R. S., Other Classes of ER/Golgi Derived Glycans (12). In *Essentials of Glycobiology*, 2 ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C.; Hart, G. W.; Etzler, M. E., Eds. Coldspring Harbor Press: Coldspring Harbor, NY, 2009; p 784.
211. Larriba, G.; Elorza, M. V.; Villanueva, J. R.; Sentandreu, R., Participation of dolichol phospho-mannose in the glycosylation of yeast wall manno-proteins at the polysomal level. *FEBS Lett.* **1976,** 71, 316–320.

212. Endo, T., *O*-mannosyl glycans in mammals. *Biochim. Biophys. Acta* **1999,** 1473, 237-246.
213. Haltiwanger, R. S.; Lowe, J. B., Role of Glycosylation in Development. *Annu. Rev. Biochem.* **2004,** 73, 491-537.
214. Finne, J.; Krusius, T.; Margolis, R. K.; Margolis, R. U., Novel mannitol-containing oligosaccharides obtained by mild alkaline borohydride treatment of a chondroitin sulfate proteoglycan from brain. *J. Biol. Chem.* **1979,** 254, 10295-10300.
215. Chiba, A.; Matsumura, K.; Yamada, H.; Inazu, T.; Shimizu, T.; Kusunoki, S.; Kanazawa, I.; Kobata, A.; Endo, T., Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve α-dystroglycan. The role of a novel O-mannosyl-type oligosaccharide in the binding of α-dystroglycan with laminin. *J. Biol. Chem.* **1997,** 272, 2156-2162.
216. Margolis, R. K.; Margolis, R. U., Nervous tissue proteoglycans. *Experientia* **1993,** 49, (5), 429–446.
217. Yuen, C. T.; Chai, W.; Loveless, R. W.; Lawson, A. M.; Margolis, R. U.; Feizi, T., Brain contains HNK-1 immunoreactive O-glycans of the sulfoglucuronyl lactosamine series that terminate in 2-linked or 2,6-linked hexose (mannose). *J. Biol. Chem.* **1997,** 272, 8924-8931.
218. Smalheiser, N. R.; Haslam, S. M.; Sutton-Smith, M.; Morris, H. R.; Dell, A., Structural analysis of sequence O-linked to mannose reveals a novel Lewis X structure in cranin (dystroglycan) purified from sheep brain. *J. Biol. Chem.* **1998,** 273, 23698-23703.
219. Williamson, R. A.; Henry, M. D.; Daniels, K. J.; Hrstka, R. F.; Lee, J. C.; Sunada, Y.; Ibraghimov-Beskrovnaya, O.; Campbell, K. P., Dystroglycan is essential for early embryonic development: disruption of Reichert's membrane in Dag1-null mice. *Hum. Mol. Genet.* **1997,** 6, 831-841.
220. Dobyns, W. B.; Pagon, R. A.; Armstrong, D.; Curry, C. J.; Greenberg, F.; Grix, A.; Holmes, L. B.; Laxova, R.; Michels, V. V.; Robinow, M.; Zimmerman, R. L.; Opitz, J. M.; Reynolds, J. F., Diagnostic criteria for Walker-Warburg syndrome. *Am. J. Med. Genet.* **1989,** 32, (2), 195–210.
221. Beltran-Valero De Bernabe, D.; Currier, S.; Steinbrecher, A.; Celli, J.; Van Beusekom, E.; Van Der Zwaag, B.; Kayserili, H.; Merlini, L.; Chitayat, D.; Dobyns, W. B.; Cormand, B.; Lehesjoki, A. E.; Cruces, J.; Voit, T.; Walsh, C. A.; van Bokhoven, H.; Brunner, H. G., Mutations in the *O*-mannosyltransferase gene POMT1 give rise to the severe neuronal migration disorder Walker-Warburg syndrome. *Am. J. Hum. Genet.* **2002,** 71, (5), 1033–1043.
222. Jurado, L. A.; Coloma, A.; Cruces, J., Identification of a human homolog of the Drosophila rotated abdomen gene (POMT1) encoding a putative protein *O*-mannosyl-transferase, and assignment to human chromosome 9q34.1. *Genomics* **1999,** 58, (2), 171–180.
223. Gentzsch, M.; Tanner, W., The PMT gene family: protein *O*-glycosylation in *Saccharomyces cerevisiae* is vital. *EMBO J.* **1996,** 15, 5752-5759.

224. Timpel, C.; Strahl-Bolsinger, S.; Ziegelbauer, K.; Ernst, J. F., Multiple functions of Pmt1p-mediated protein *O*-mannosylation in the fungal pathogen *Candida albicans*. *J. Biol. Chem.* **1998,** 273, 20837-20846.
225. Girrbach, V.; Zeller, T.; Priesmeier, M.; Strahl-Bolsinger, S., Structure-Function Analysis of the Dolichyl Phosphate-Mannose: Protein O-Mannosyltransferase ScPmt1p*. *J. Biol. Chem.* **2000,** 275, (25), 19288-19298.
226. Strahl-Bolsinger, S.; Gentzsch, M.; Tanner, W., Protein *O*-mannosylation. *Biochim. Biophys. Acta* **1999,** 1426, (2), 297-307.
227. Furmanek, A., and Hofsteenge, J., Protein C-mannosylation: Facts and questions. *Acta Biochemica Polonica* **2000,** 47, (3), 781-789.
228. Krieg, J., Glasner, W., Vicentini, A., Doucey, M.-A., Loffler, A., Hess, D., and Hofsteenge, J., C-Mannosylation of Human RNase 2 Is an Intracellular Process Performed by a Variety of Cultured Cells. *J. Biol. Chem.* **1997,** 272, (42), 26687-26692.
229. Hofsteenge, J.; Müller, D. R.; de Beer, T.; Löffler, A.; Richter, W. J.; Vliegenthart, J. F., New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. *Biochemistry* **1994,** 33, 13524-13530.
230. Falzarano, D., Krokhin, O., Van Domselaar, G., Wolf, K., Seebach, J., Schnittler, H.-J., and Feldmann, H., Ebola sGP The first viral glycoprotein shown to be C-mannosylated. *Virology* **2007,** 368, 83-90.
231. Fahey, S. J.; Carroll, A. R., Natural Products Isolated From Species of Halgerda Bergh, 1880 (Mollusca: Nudibranchia) and Their Ecological and Evolutionary Implications. *J. Chem. Ecol.* **2007,** 33, 1226-1234.
232. Vliegenthart, J. F. G., and Casset, F., Novel forms of protein glycosylation. *Curr. Opin. in Struct. Biol.* **1998,** 8, 565-571.
233. Hofsteenge, J.; Blommers, M.; Hess, D.; Furmanek, A.; Miroshnichenko, O., The Four Terminal Components of the Complement System Are C-Mannosylated on Multiple Tryptophan Residues. *J. Biol. Chem.* **1999,** 274, (46), 32786-32794.
234. Furmanek, A.; Hofsteenge, J., Protein C-mannosylation: Facts and questions. *Acta Biochimica Polonica* **2000,** 47, (3), 781-789.
235. Nishikawa, T.; Kajii, S.; Sato, C.; Yasukawa, Z.; Kitajimac, K.; Isobea, M., α-C-Mannosyltryptophan is not recognized by conventional mannose-binding lectins. *Bioorg. and Med. Chem.* **2004,** 12, 2343–2348.
236. Raju, T. S.; Briggs, J. B.; Chamow, S. M.; Winkler, M. E.; Jones, A. J., Glycoengineering of therapeutic glycoproteins: in vitro galactosylation and sialylation of glycoproteins with terminal N-acetylglucosamine and galactose residues. *Biochemistry* **2001,** 40, 8868-8876.
237. Elliott, S.; Lorenzini, T.; Asher, S.; Aoki, K.; Brankow, D.; Buck, L.; Busse, L.; Chang, D.; Fuller, J.; Grant, J.; Hernday, N.; Hokum, M.; Hu, S.; Knudten, A.; Levin, N.; Komorowski, R.; Martin, F.; Navarro, R.; Osslund, T.; Rogers, G.; Rogers, N.; Trail. G.; Egrie, J., Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotechnol.* **2003,** 21, 414-421.

238. Wieland, F., Structure and biosynthesis of prokaryotic glycoproteins. *Biochimie* **1988,** 70, 1493–1504.
239. Schirm, M.; Kalmokoff, M.; Aubry, A.; Thibault, P.; Sandoz, M.; Logan, S. M., Flagellin from Listeria monocytogenes is glycosylated with β-O-linked N-acetylglucosamine. *J. Bacteriol.* **2004,** 186, 6721–6727.
240. Mendez, C., and Salas, J., Altering the glycosylation pattern of bioactive compounds. *Trends Biotech.* **2001**, (19), 449-456.
241. He, X. m.; Liu, H. w., Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annu Rev Biochem* **2002,** 71, 701-54.
242. Schobert, R.; Schlenk, A., Tetramic and tetronic acids: An update on new derivatives and biological aspects. *Bioorg. Med. Chem.* **2008,** 16, (6), 4203-4221.
243. Waitz, J. A.; Horan, A. C.; Kalyanpur, M.; Lee, B. K.; Loebenberg, D.; Marquez, J. A.; Miller, G.; Patel, M. G., Kijanimicin (Sch 25663), a novel antibiotic produced by *Actinomadura kijaniata* SCC 1256. Fermentation, isolation, characterization and biological properties. *J. Antibiot.* **1981,** 34, (8), 1101-1106.
244. Mallams, A. K.; Puar, M. S.; Rossman, R. R., Kijanimicin 1. Structures of the Individual Sugar Components. *J. Am. Chem. Soc.* **1981,** 103, 3938-3940.
245. Mallams, A. K.; Puar, M. S.; Rossman, R. R.; McPhail, A. T.; Macfarlane, R. D., Kijanimicin 2. Structure and Absolute Stereochemistry of Kijanimicin. *J. Am. Chem. Soc.* **1981,** 103, 3940-3943.
246. Mallams, A. K.; Puar, M. S.; Rossman, R. R.; McPhail, A. T.; Macfarlane, R. D.; Stephens, R. L., Kijanimicin 3. Structure and Absolute Stereochemistry of Kijanimicin. *J. Am. Chem. Soc. Perkin Trans.* **1983,** 1, 1497-1534.
247. Wang, L.; White, R.; Vining, L. C., Biosynthesis of the dideoxysugar component of jadomycin B: genes in the jad cluster of Streptomyces venezuelae ISP5230 for L-digitoxose assembly and transfer to the angucycline aglycone. *Microbiology* **2002,** 148, 1091-1103.
248. Zielinski, J.; Jereczek, E.; Sowinski, P.; Falkowski, L.; Rudowski, A.; Borowski, E., The structure of a novel sugar component of polyene macrolide antibiotics: 2,6-dideoxy-l-ribohexopyranose. *J. Antibiot.* **1979,** 32, 565-568.
249. Tomita, F.; Tamaoki, T., Tetrocarcins, novel antitumor antibiotics. I. Producing organism, fermentation and antimicrobial activity. *J. Antibiot.* **1980,** 33, (9), 940-945.
250. Rezanka, T.; Dembitsky, V. M., Ten-Membered Substituted Cyclic 2-Oxecanone (Decalactone) Derivatives from Latrunculia corticata, a Red Sea Sponge. *Eur. J. Org. Chem* **2003,** 2003, (11), 2144-2152.
251. Kiliani, H., Ueber Digitoxin. *Archiv. der Pharm.* **1896**, 481-489.
252. Kanchanapoom, T.; Kasai, R.; Ohtani, K.; Andriantsiferana, M.; Yamasaki, K., Pregnane and Pregnane Glycosides from the Malagasy Plant, Cynanchum aphyllum. *Chem. Pharm. Bull.* **2002,** 50, 1031-1034.
253. Murakami, R.; Tomikawa, T.; Shin-Ya, K.; Shinozaki, J.; Kajiura, T.; Seto, H.; Hayakawa, Y., Ammocidin, a new apoptosis inducer in Ras-dependent cells from Saccharothrix sp. II. Physicochemical properties and structure elucidation. *J. Antibiot.* **2001,** 54, 714-717.

254. Kong, F.; Zhao, N.; Siegel, M.; Janota, K.; Ashcroft, J. S.; Koehn, F. E.; Borders, D. B.; Carter, G. T., Saccharomycins, Novel Heptadecaglycoside Antibiotics Effective Against Multidrug-Resistant Bacteria. *J. Am. Chem. Soc.* **1998,** 120, (51), 13301-13311.
255. Fischer, C.; Rodriguez, L.; Patallo, E. P.; Lipata, F.; Brana, A. F.; Mendez, C.; Salas, J. A.; Rohr, J., Digitoxosyltetracenomycin C and Glucosyltetracenomycin C, Two Novel Elloramycin Analogues Obtained by Exploring the Sugar Donor Substrate Specificity of Glycosyltransferase ElmGT. *J. Nat. Prod.* **2002,** 65, 1685-1689.
256. Lombo, F.; Gibson, M.; Greenwall, L.; Brana, A. F.; Rohr, J.; Salas, J. A.; Mendez, C., Engineering Biosynthetic Pathways for Deoxysugars: Branched-Chain Sugar Pathways and Derivatives from the Antitumor Tetracenomycin. *Chem. Biol.* **2004,** 11, 1709-1718.
257. Brimacombe, J. S.; Hanna, R.; Saeed, M. S.; Tucker, L. C. N., Convenient Syntheses of L-Digitoxose, L-Cymarose, and L-Ristosamine. *J. Chem. Soc. Perkin Trans.* **1982,** I, 2583-2587.
258. Fuchs, E. F.; Horton, D.; Weckerle, W.; Winter-Mihaly, E., Synthesis and antitumor activity of sugar-ring hydroxyl analogs of daunorubicin. *J. Med. Chem.* **1979,** 22, (4), 406-411.
259. Roush, W. R.; Brown, R. J., Total Synthesis of Carbohydrates: Stereoselective Syntheses of 2,6-Dideoxy-D-arabino-hexose and 2,6-Dideoxy-D-ribohexose. *J. Org. Chem.* **1982,** 47, (7), 1373-1378.
260. Fronza, G.; Fuganti, C.; Graselli, P.; Pedrocchi-Fantoni, G.; Zirotti, C., On the Steric Course of the Addition of Diallyl zinc onto α,ß-Dialkoxy Chiral Carbonyl Compounds Stereospecific Synthesis of 2,6-Dideoxysugars of the L-series. *Tet. Lett.* **1982,** 23, (40), 4143-4146.
261. Braun, M.; Moritz, J., A Convenient Route to Both Enantiomers of Digitoxose. *Syn. Lett.* **1991**, 750-752.
262. Zhu, L.; Kedenburg, J. P.; Xian, M.; Wang, G. P., A systematic strategy for preparation of uncommon sugars through enzymatic resolution and ring-closing metathesis. *Tet. Lett.* **2005,** 46, 811-813.
263. Bock, K.; Lundt, I.; Pederson, C., The Base Catalyzed Rearrangement of Some 6-Bromo-2,6-dideoxyaldono-1,4-lactones. Preparation of L-Digitoxose. *Acta. Chem. Scand.* **1984,** 38B, 555-561.
264. Moffatt, J. G.; Khorana, H. G., The Total Synthesis of Coenzyme A. *J. Am. Chem. Soc.* **1959,** 81, (5), 1265.
265. Madduri, K.; Waldron, C.; Merlo, D. J., Rhamnose biosynthesis pathway supplies precursors for primary and secondary metabolism in *Saccharopolyspora spinosa*. *J. Bacteriol.* **2001,** 183, (19), 5632-5638.
266. Ikeda, H.; Nonomiya, T.; Usami, M.; Ohta, T.; Omura, S., Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Nat. Acad. Sci. USA* **1999,** 96, (17), 9509-9514.
267. González, A.; Remsing, L. L.; Lombó, F.; Fernández, M. J.; Prado, L.; Braña, A. F.; Künzel, E.; Rohr, J.; Méndez, C.; Salas, J. A., The mtmVUC genes of the

mithramycin gene cluster in *Streptomyces argillaceus* are involved in the biosynthesis of the sugar moieties. *Mol. Gen. Genet.* **2001,** 264, (6), 827-835.
268. Draeger, G.; Park, S.-H.; Floss, H. G., Mechanism of the C-2 deoxygenation step in the biosynthesis of the deoxyhexose moieties of the antibiotics granaticin and oleandomycin. *J. Am. Chem. Soc.* **1999,** 121, (11), 2611-2612.
269. Bate, N.; Butler, A. R.; Smith, I. P.; Cundliffe, E., The mycarose-biosynthetic genes of *Streptomyces fradiae*, producer of tylosin. *Microbiol.* **2006,** 146, 139-146.
270. Bihlmaier, C.; Welle, E.; Hofmann, C.; Welzel, K.; Vente, A.; Breitling, E.; Müller, M.; Glaser, S.; Bechthold, A., Biosynthetic Gene Cluster for the Polyenoyltetramic Acid α-Lipomycin. *Antimicrob. Agents Chemother.* **2006,** 50, (6), 2113-2121.
271. Chen, H.; Agnihotri, G.; Guo, Z.; Que, N. L. S.; Chen, X.; Liu, H.-w., Biosynthesis of mycarose: isolation and characterization of enzymes involved in the C-2 deoxygenation. *J. Am. Chem. Soc.* **1999,** 121, 8124-8125.
272. Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976,** 72, 248-254.
273. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** 227, (5259), 680-685.
274. Sambrook, J.; Russell, D. W., *Molecular Cloning: A Laboratory Manual*. 3rd ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY: 2001.
275. Dairi, T.; Hamano, Y.; Furumai, T.; Oki, T., Development of a self-cloning system for *Actinomadura verrucosospora* and identification of polyketide synthase genes essential for production of the angucyclic antibiotic pradimicin. *Appl. Env. Micro.* **1999,** 65, (6), 2703-2709.
276. Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A., *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich, England: 2000.
277. Birnboim, H. C.; Doly, J., A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **1979,** 7, (6), 1513-1523.
278. Vesterberg, O., Staining of Protein Zones after Isoelectric Focusing in Polyacrylamide Gels. *Biochim. Biophys. Acta.* **1971,** 243, 345-348.
279. Burnette, W. N., Western Blotting: Electrophoretic Transfer of Proteins from Sodium Dodecyl Sulfate-Polyacrylamide Gels to Unmodified Nitrocellulose and Radiographic Detection with Antibody and radioiodinated Protein A. *Anal. Biochem.* **1981,** 112, 195-203.
280. Takahashi, H., Liu, Y-n., Liu, H-w., A two-stage one-pot enzymatic synthesis of TDP-L-mycarose from thymidine and glucose-1-phosphate. *Journal of the American Chemical Society* **2006,** 128, 1432-1433.
281. Blakeney, A. B.; Harris, P. J.; Henry, R. J.; Stone, B. A., A simple and rapid preparation of alditol acetates for monosaccharide analysis. *Carbohydr. Res.* **1983,** 113, 291-299.

282. Weigel, T. M.; Liu, L.-d.; Liu, H.-w., Mechanistic Studies of the Biosynthesis of 3,6-Dideoxyhexoses in *Yersinia pseudotuberculosis*: Purification and Characterization of CDP-4-Keto-6-deoxy-D-glucose-3 -dehydrase. *Biochemistry* **1992,** 31, (7), 2129-2139.
283. Glaser, L.; Kornfeld, S., The Enzymatic Synthesis of Thymidine-linked Sugars II. THYMIDINE DIPHOSPHATE L-RHAMNOSE. *J. Biol. Chem.* **1961,** 236, (6), 1795-1799.
284. Kornfeld, S.; Glaser, L., The Enzymatic Synthesis of Thymidine-linked Sugars I. THYMIDINE DIPHOSPHATE GLUCOSE. *J. Biol. Chem.* **1961,** 236, (6), 1791-1794.
285. Thibodeaux, C. J.; Melancon, C. E.; Liu, H.-w., Natural product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed. Engl.* **2008,** 47, 9814-9859.
286. Hutchinson CR, F. I., Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu. Rev. Microbiol.* **1995,** 49, 201-238.
287. Hong, L. Biochemical studies of the enzymes involved in deoxysugar D-forosamine biosynthesis. University of Texas, Austin, 2004.
288. Amann, S.; Drager, G.; Rupprath, G.; Kirschning, A.; Ellinga, L., (Chemo)enzymatic synthesis of dTDP-activated 2,6-dideoxysugars as building blocks of polyketide antibiotics. *Carbohydr. Res.* **2001,** 335, 23-32.
289. Neumann, T.; Heinze, S.; Ritzau, M.; Martin, K.; Moellmann, U.; Reiche, R.; Hentschel, E. J. New macrolide antibiotic from *Paenibacillus larvae*, method for its production, and its use. 1998.
290. Kamigoori, K.; Kawashima, A.; Akama, T.; Oota, M.; Tamai, M.; Hanada, K. Hypolipemics containing chlorothricin. 1990.
291. A., L.; Bechthold, A., It works: combinatorial biosynthesis for generating novel glycosylated compounds. *Mol. Microbiol.* **2005,** 58, (1), 3-5.
292. Luzhetskyy A.; Bechthold, A., Features and applications of bacterial glycosyltransferases: current state and prospects. *Appl. Microbiol. Biotechnol.* **2008,** 80, 945-952.
293. Jia, X. Y.; Tian, Z. H.; Shao, L.; Qu, X. D.; Zhao, Q. F.; Tang, J.; Tang, G. L.; Liu, W., Genetic characterization of the chlorothricin gene cluster as a model for spirotetronate antibiotic biosynthesis. *Chem Biol* **2006,** 13, (6), 575-85.
294. Jiang, Z.-D.; Jensen, P. R.; Fenical, W., Lobophorins A and B, new antiinflammatory macrolides produced by a tropical marine bacterium. *Bioorg. & Med. Chem. Lett.* **1999,** 9, (14), 2003-2006.
295. Igarashi, Y.; Takagi, K.; Kan, Y.; Fujii, K.; Harada, K.-I.; Furumai, T.; Oki, T., Arisostatins A and B, new members of tetrocarcin class of antibiotics from *Micromonospora sp*. TP-A0316. II. Structure determination. *J. Antibiot.* **2000,** 53, (3), 233-240.
296. Kaneko, M.; Nakashima, T.; Uosaki, Y.; Hara, M.; Ikeda, S.; Kanda, Y., Synthesis of tetrocarcin derivatives with specific inhibitory activity towards Bcl-2 functions. *Bioorg. & Med. Chem. Lett.* **2001,** 11, (7), 887-890.

297. Luzhetskyy, A.; Bechthold, A., Features and applications of bacterial glycosyltransferases: current state and prospects. *Appl. Microbiol. Biotechnol.* **2008,** 80, 945-952.
298. Luzhetskyy, A.; Fedoryshyn, M.; Dürr, C.; Taguchi, T.; Novikov, V.; Bechthold, A., Iteratively acting glycosyltransferases involved in the hexasaccharide biosynthesis of landomycin A. *Chem. Biol.* **2005,** 12, 725–729.
299. Zhang, C.; Griffith, B. R.; Fu, Q.; Albermann, C.; Fu, X.; Lee, I. K.; Li, L.; Thorson, J. S., Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **2006,** 313, 1291-1294.
300. Thiem, J., Kopper, S., Syntheses of Kijanimcin Oligosaccharides. *Tetrahedron* **1990,** 46, (1), 113-138.
301. Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E., Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp. Gene* **1992,** 116, (1), 43-49.
302. Wohlert, S.; Lomovskaya, N.; Kulowski, K.; Fonstein, L.; Occi, J. L.; Gewain, K. M.; MacNeil, D. J.; Hutchinson, C. R., Insights about the biosynthesis of the avermectin deoxysugar L-oleandrose through heterologous expression of Streptomyces avermitilis deoxysugar genes in Streptomyces lividans. *Chem. Biol.* **2001,** 8, (681-700).
303. Vara, J.; Lewandowska-Skarbek, M.; Wang, Y.-G.; Donadio, S.; Hutchinson, C. R., Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in saccharopolyspora erythraea (Streptomyces erythreus). *J. Bacteriol.* **1989,** 171, 5872–5881.
304. Bibb, M. J.; Sherman, D. H.; Omura, S.; Hopwood, D. A., Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of a polyketide antibiotic frenolicin. *Gene* **1994,** 142, 31-39.
305. Kwon, H.-j.; Smith, W. C.; Scharon, A. J.; Hwang, S. H.; Kurth, M. J.; Shen, B., C-O Bond Formation by Polyketide Synthases. *Science* **2002,** 297, (5585), 1327-1330.
306. Takahashi, H.; Liu, Y.-n.; Liu, H.-w., A two-stage one-pot enzymatic synthesis of TDP-L-mycarose from thymidine and glucose-1-phosphate. *J. Am. Chem. Soc.* **2006,** 128, 1432-1433.
307. von Mulert, U.; Luzhetskyy, A.; Hofmann, C.; Mayer, A.; Bechthold, A., Expression of the landomycin biosynthetic gene cluster in a PKS mutant of Streptomyces fradiae is dependent on the coexpression of a putative transcriptional activator gene. *FEMS Microbiol. Lett.* **2004,** 230, (1), 91–97.
308. Troutman, J. M.; Imperiali, B., Campylobacter jejuni PglH Is a Single Active Site Processive Polymerase that Utilizes Product Inhibition to Limit Sequential Glycosyl Transfer Reactions. *Biochemistry* **2009,** 48, (12), 2807-2816.
309. Zhang, C.; Albermann, C.; Fu, X.; Thorson, J. S., The in vitro characterization of the iterative avermectin glycosyltransferase AveBI reveals reaction reversibility and sugar nucleotide flexibility. *J. Am. Chem. Soc.* **2006,** 128, 16420–16421.
310. Minami, A.; Kakinuma, K.; Eguchi, T., Aglycon switch approach toward unnatural glycosides from natural glycoside with glycosyltransferase VinC. *Tet. Lett.* **2005,** 46, (37), 6187-6190.

311. Zhao, L.; Beyer, N. J.; Borisova, S. A.; Liu, H.-w., *β*-Glucosylation as a Part of Self-Resistance Mechanism in Methymycin/Pikromycin Producing Strain *Streptomyces venezuelae*. *Biochemistry* **2003,** 42, (50), 14794-14804.
312. Quiros, L. M.; Aquirrezabalaga, I.; Olano, C.; Méndez, C.; Salas, J. A., Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by Streptomyces antibioticus. *Mol. Microbiol.* **1998,** 28, (6), 1177−1185.
313. de Beer, T., Vliegenthart, J. F. G., Loffler, A., and Hofsteenge, J., The Hexopyranosyl Residue That Is C-Glycosidically Linked to the Side Chain of Tryptophan-7 in Human RNase Us Is a-Mannopyranose. *Biochemistry* **1995,** 34, (37), 11785-11789.
314. Krieg, J., Hartmann, S., Vicentini, A., Glasner, W., Hess, D., and Hofsteenge, J., Recognition Signal for C-Mannosylation of Trp-7 in RNase 2 Consists of Sequence Trp-x-x-Trp. *Mol. Biol. Cell* **1998,** 9, 301-309.
315. Doucey, M.-A., Hess, D., Cacan, R., and Hofsteenge, J., Protein C-Mannosylation Is Enzyme-catalysed and Uses Dolichyl-Phosphate-Mannose as a Precursor. *Mol. Biol. Cell* **1998,** 9, 291-300.
316. Fahey, S. J., and Carroll, A. R., Natural Products Isolated From Species of Halgerda Bergh, 1880 (Mollusca: Nudibranchia) and Their Ecological and Evolutionary Implications. *J. Chem. Ecol.* **2007,** 33, 1226-1234.
317. Hartmann, S.; Hofsteenge, J., Properdin, the positive regulator of the complement, is highly C-mannosylated. *J. Biol. Chem.* **2000,** 275, 28569-28574.
318. Hofsteenge, J.; Blommers, M.; Hess, D.; Furmanek, A.; Miroshnichenko, O., The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. *J. Biol. Chem.* **1999,** 274, 32786-32794.
319. Leonidas, D.; Boix, E.; Prill, R.; Suzuki, M.; Turton, R.; Minson, K.; Swaminathan, J.; Youle, R.; Acharya, K., Mapping the Ribonucleolytic Active Site of Eosinophil-derived Neurotoxin (EDN). *J. Biol. Chem.* **2001,** 276, 15009-15017.
320. Hilton, D. J., et al., *J. Biol. Chem.* **1996,** 271, 4699-4708.
321. Furmanek, A., Hess, D., Rogniaux, H., and Hofsteenge, J., The WSAWS Motif Is C-Hexosylated in a Soluble Form of the Erythropoietin Receptor. *Biochemistry* **2003,** 42, (28), 8452-8458.
322. Wildt, S.; Gerngross, T., The Humanization of N-Glycosylation Pathways in Yeast. *Nature Rev. Microbio.* **2005,** 3, 119-128.
323. Pastores, G., In NIH/University of Washington: Seattle, 2005.
324. Barranger, J. A.; O'Rourke, E., Lessons learned from the development of enzyme therapy for Gaucher disease. *J. Inherit. Metab. Dis.* **2001,** 24, (Suppl. 2), 122-126.
325. Sinclair, A. M.; Elliott, S., Glycoengineering: The Effect of Glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci.* **2005**, (8), 1625-1635.
326. Ihara, Y.; Manabe, S.; Kanda, M.; Kawano, H.; Nakayama, T.; Sekine, I.; Kondo, T.; Ito, Y., Increased expression of protein C-mannosylation in the aortic vessels of diabetic Zucker rats. *Glycobiology* **2004,** 15, (4), 383-392.

327. Ervin, L. A.; Ball, L. E.; Crouch, R. K.; Schey, K. L., Phosphorylation and Glycosylation of Bovine Lens MP20. *Investigative Ophthalmology & Visual Science* **2005,** 46, (2), 627-635.
328. Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M., Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical Chemistry* 68, 850-858.
329. Graham, J., *Centrifugation; a practical approach.* IRL Press: Oxford, 1985; p 161-182.
330. Tao, H.; Bausch, C.; Richmond, C.; Blattner, F. R.; Conway, T., Functional Genomics: Expression Analysis of Escherichia coli Growing on Minimal and Rich Media. *J. Bacteriol.* **1999,** 181, (20), 6425-6440.
331. Schutzbach, J. S.; Zimmerman, J. W.; Forsee, W. T., The Purification and Characterization of Recombinant Yeast Dolichyl-phosphate-mannose Synthase. *J. Biol. Chem.* **1993,** 268, 24190-24196.
332. Kozak, M., An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **1987,** 15, (20), 8125-8148.
333. Gao, X.; Yo, P.; Keith. A.; Ragan, T. J.; Harris, T. K., Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **2003,** 31, (22), e143.
334. Kjeldsen T. et al., Prepro-Leaders Lacking N-Linked Glycosylation for Secretory Expression in the Yeast Saccharomyces cerevisiae. *Protein Expression and Purification* **1998,** 14, 309-316.
335. Schmitt, M. J., Cloning and expression of a cDNA copy of the viral K28 killer toxin gene in yeast. *Mol Gen Genet.* **1995,** 246, (2), 236-246.
336. Yaffe, M. P.; Schatz, G., Two nuclear mutations that block mitochondrial protein import in yeast. *Proc. Nat. Acad. Sci. USA* **1984,** 81, (15), 4819-4823.
337. Fernandez, et al., *Glycobiology* **2002,** 12, 555-562.
338. Shimoni, Y.; Schekman, R., Vesicle budding from Endoplasmic Reticulum. *Methods Enzymol.* **2002,** 351, 258-278.
339. Sanyal, S.; Frank, C. G.; Menon, A. K., Distinct Flippases Trasnlocate Glycerophospholipids and Oligosaccharide Diphosphate Dolichols across the Endoplasmic Reticulum. *Biochemistry* **2008,** 47, 7937-7946.
340. Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G., Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Nat. Acad. Sci. USA* **2004,** 101, (24), 8882-8887.
341. Burdine, L.; Gilette, T. G.; Lin, H.-J.; Kodadek, T., Periodate-Triggered Cross-Linking of DOPA-Containing Peptide−Protein Complexes. *J. Am. Chem. Soc.* **2004,** 126, (37), 11442-11443.
342. Liu, B.; Burdine, L.; Kodadek, T., Chemistry of periodate-mediated cross-linking of 3,4-dihydroxylphenylalanine-containing molecules to proteins. *J. Am. Chem. Soc.* **2006,** 128, (47), 15228-35.

343. Oriol, R.; Martinez-Duncker, I.; Chantret, I.; Mollicone, R.; Codogno, P., Common Origin and Evolution of Glycosyltransferases Using Dol-P-monosaccharides as Donor Substrate. *Mol. Biol. Evol.* **2002,** 19, (9), 1451-1463.
344. Ponting, C. P., Novel repeats in ryanodine and IP3 receptors and protein O -mannosyltransferases. *TIBS* **2000,** 25, 48-50.
345. Lehle, L., Strahl, S., and Tanner, W., Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angew. Chem. Int. Ed.* **2006,** 45, 6802-6818.
346. Mehrle, A.; Rosenfelder, H., LIFEdb. In Molecular Genome Analysis (DKFZ) and SMP-Cell (SMP-Cell): 2004.
347. Miroux, B.; Walker, J., Over-production of Proteins in Escherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **1996,** 260, 289-298.
348. Doucey, M.-A., Hess, D., Blommers, M. J. J., and Hofsteenge, J., Recombinant human interleukin-12 is the second example of a C-mannosylated protein. *Glycobiology* **1999,** 9, (5), 435-441.
349. Freigassner, M.; Pichler, H.; Gleider, A., Tuning microbial hosts for membrane protein production. *Microbial Cell Factories* **2009,** 8, (69).
350. Rademacher, T. W., Parekh, R. B., and Dwek, R. A., Glycobiology. *Ann. Rev. Biochem.* **1988,** 57, 785-838.
351. Varki, A.; Marth, J., Oligosaccharides in vertebrate development. *Seminars in Developmental Biology* **1995,** 6, 127-138.
352. Jaeken, J.; Matthijs, G., Congenital Disorders of Glycosylation: A Rapidly Expanding Disease Family. *Annu. Rev. Genomics Hum. Genet.* **2007,** 8, 261-278.
353. Bridges, C. B.; Morgan, T. H., *The Third-Chromosome Group of Mutant Characters of Drosophila melanogaster.* Carnegie Institute: Washington, DC, 1923; p 251.
354. Shafi, R., et al., The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl. Acad. Sci. USA* **2000,** 97, 5735-5739.
355. Comer, F. I.; Hart, G. W., *O*-Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* **2000,** 275, (38), 29179-29182.
356. Love, D. C.; Hanover, J. A., The Hexosamine Signaling Pathway: Deciphering the "O-GlcNAc Code". *Science Signalling* **2005,** 312, 1-14.
357. http://www.ncbi.nlm.gov
358. Doucey, M.-A.; Hess, D.; Blommers, M. J. J.; Hofsteenge, J., Recombinant human interleukin-12 is the second example of a C-mannosylated protein. *Glycobiology* **1999,** 9, (5), 435-441.
359. Takada, A.; Kawaoka, Y., The pathogenesis of Ebola hemorrhagic fever. *Trends Microbiol.* **2001,** 9, (10), 506-511.
360. Zitvogel, L.; Lotze, M. T., Role of interleukin-12 (IL12) as an anti-tumour agent: experimental biology and clinical applications. *Res. Immunol.* **1995,** 146, 628-638.

361. Falzarano, D.; Krokhin, O.; Wahl-Jensen, V.; Seebach, J.; Wolf, K.; Schnittler, H.-J.; Feldmann, H., Structure-Function Analysis of the Soluble Glycoprotein, sGP, of Ebola Virus. *ChemBioChem* **2006,** 7, 1605-1611.
362. Barrientos, L. G.; Martin, A. M.; Wohlhueter, R. M.; Rollin, P. E., Secreted Glycoprotein from Live Zaire ebolavirus– Infected Cultures: Preparation, Structural and Biophysical Characterization, and Thermodynamic Stability. *J. Infect. Dis.* **2007,** 196, (Suppl 2), 220-231.
363. Lynn, D. E., Methods for Maintaining Insect Cell Cultures. *Journal of Insect Science* **2002,** 2.9, 1-6.
364. Tomazin, R.; Hill, A. B.; Jugovic, P.; York, I.; van Endert, P.; Ploegh, H. L.; Andrews, D. W.; Johnson, D. C., Stable binding of the herpes simplex virus ICP47 protein to the peptide binding site of TAP. *EMBO J.* **1996,** 15, (13), 3256-3266.
365. Jennewein, S.; Rithner, C. D.; Williams, R. M.; Croteau, R. B., Taxol biosynthesis: Taxane 13α-hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc. Nat. Acad. Sci. USA* **2001,** 98, (24), 13595-13600.
366. Chaney, W.; Sundaram, S.; Friedman, N.; Stanley, P., The Lec4A CHO Glycosylation Mutant Arises from Miscompartmentalization of a Golgi Glycosyltranferase. *J. Cell. Biol.* **1989,** 109, 2089-2096.
367. O' Reilly, D.; Miller, L. K.; Luckow, V. A., *Baculovirus Expression Vectors*. Oxford University Press: New York, 1994.
368. Hofsteenge, J., Huweiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F., and Peter-Katalinic, J., C-Mannosylation and O-Fucosylation of the Thrombospondin Type 1 Module. *J. Biol. Chem.* **2001,** 276, (9), 6485–6498.
369. Laughlin, S.; Bertozzi, C., Imaging the Glycome. *Proc. Nat. Acad. Sci. USA* **2009,** 106, (1), 12-17.
370. Loffler, A., Doucey, M.-A., Jansson. A. M., Muller, D. R., de Beer, T., Hess, D., Meldal, M., Richter. W. J., Vliegenthart, J. F. G., and Hofsteenge, J., Spectroscopic and Protein Chemical Analyses Demonstrate the Presence of C-Mannosylated Tryptophan in Intact Human RNase 2 and Its Isoforms. *Biochemistry* **1996,** 35, (37), 12005-12014.
371. Zanetta, J.-P.; Pons, A.; Richet, C.; Huet, G.; Timmerman, P.; Leroy, Y.; Bohin, A.; Bohin, J.-P.; Trinel, P.-A.; Poulain, D.; Hofsteenge, J., Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins. *Anal. Biochem.* **2004,** 329, 199–206.